(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10387 A2

(51) International Patent Classification[7]: C12N 15/12, C07K 14/705, 16/28, C12Q 1/68, G01N 33/50, 33/577, A61K 38/17, A01K 67/027

(21) International Application Number: PCT/US01/23433

(22) International Filing Date: 25 July 2001 (25.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/221,478	27 July 2000 (27.07.2000)	US
60/223,268	3 August 2000 (03.08.2000)	US
60/227,054	21 August 2000 (21.08.2000)	US
60/231,121	8 September 2000 (08.09.2000)	US
60/232,243	13 September 2000 (13.09.2000)	US
60/232,691	15 September 2000 (15.09.2000)	US
60/235,146	22 September 2000 (22.09.2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way, #1, Menlo Park, CA 94025 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Redwood City, CA 94025 (US). **ELLIOT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **KALLICK, Deborah, A.** [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **POLICKY, Jennifer, L.** [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). **KEARNEY, Liam** [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). **GRAUL, Richard, C.** [—/US]; 682-29th Avenue, San Francisco, CA 94121 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parwood Drive #2, Cupertino, CA 95014 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **DING, Li** [US/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. The invention also provides expression vectors, host cells, antibodies, agonist, and antagonist. The invention also provides mehtods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.

G-PROTEIN COUPLED RECEPTORS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of G-protein coupled receptors and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals. Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus activated, triggers an intracellular biochemical cascade that ends with the activation of an intracellular target molecule, such as a transcription factor. This process of signal transduction regulates all types of cell functions including cell proliferation, differentiation, and gene transcription. The G-protein coupled receptors (GPCRs), encoded by one of the largest families of genes yet identified, play a central role in the transduction of extracellular signals across the plasma membrane. GPCRs have a proven history of being successful therapeutic targets.

GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic transmembrane domains which together form a bundle of antiparallel alpha (α) helices. GPCRs range in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10; Coughlin, S.R. (1994) Curr. Opin. Cell Biol. 6:191-197). The amino-terminus of a GPCR is extracellular, is of variable length, and is often glycosylated. The carboxy-terminus is cytoplasmic and generally phosphorylated. Extracellular loops alternate with intracellular loops and link the transmembrane domains. Cysteine disulfide bridges linking the second and third extracellular loops may interact with agonists and antagonists. The most conserved domains of GPCRs are the transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account, in part, for structural and functional features of the receptor. In most cases, the bundle of α helices forms a ligand-binding pocket. The extracellular N-terminal segment, or one or more of the three extracellular loops, may also participate in ligand binding. Ligand binding activates the receptor by inducing a conformational change in intracellular portions of the receptor. In turn, the large, third intracellular loop of the activated receptor interacts with a heterotrimeric guanine nucleotide binding (G) protein complex which mediates further intracellular signaling activities, including the activation of second messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and

the interaction of the activated GPCR with ion channel proteins. (See, e.g., Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6; Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176; Baldwin, J.M. (1994) Curr. Opin. Cell Biol. 6:180-190.)

GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine, γ-aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropic-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli that have yet to be identified are known as orphan receptors.

The diversity of the GPCR family is further increased by alternative splicing. Many GPCR genes contain introns, and there are currently over 30 such receptors for which splice variants have been identified. The largest number of variations are at the protein C-terminus. N-terminal and cytoplasmic loop variants are also frequent, while variants in the extracellular loops or transmembrane domains are less common. Some receptors have more than one site at which variance can occur. The splicing variants appear to be functionally distinct, based upon observed differences in distribution, signaling, coupling, regulation, and ligand binding profiles (Kilpatrick, G.J. et al. (1999) Trends Pharmacol. Sci. 20:294-301).

GPCRs can be divided into three major subfamilies: the rhodopsin-like, secretin-like, and metabotropic glutamate receptor subfamilies. Members of these GPCR subfamilies share similar functions and the characteristic seven transmembrane structure, but have divergent amino acid sequences. The largest family consists of the rhodopsin-like GPCRs, which transmit diverse extracellular signals including hormones, neurotransmitters, and light. Rhodopsin is a photosensitive GPCR found in animal retinas. In vertebrates, rhodopsin molecules are embedded in membranous stacks found in photoreceptor (rod) cells. Each rhodopsin molecule responds to a photon of light by triggering a decrease in cGMP levels which leads to the closure of plasma membrane sodium channels. In this manner, a visual signal is converted to a neural impulse. Other rhodopsin-like GPCRs are directly involved in responding to neurotransmitters. These GPCRs include the receptors for adrenaline (adrenergic receptors), acetylcholine (muscarinic receptors), adenosine, galanin, and glutamate (N-methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkinstall (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 7-9, 19-22,

32-35, 130-131, 214-216, 221-222; Habert-Ortoli, E. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9780-9783.)

The galanin receptors mediate the activity of the neuroendocrine peptide galanin, which inhibits secretion of insulin, acetylcholine, serotonin and noradrenaline, and stimulates prolactin and growth hormone release. Galanin receptors are involved in feeding disorders, pain, depression, and Alzheimer's disease (Kask, K. et al. (1997) Life Sci. 60:1523-1533). Other nervous system rhodopsin-like GPCRs include a growing family of receptors for lysophosphatidic acid and other lysophospholipids, which appear to have roles in development and neuropathology (Chun, J. et al. (1999) Cell Biochem. Biophys. 30:213-242).

The largest subfamily of GPCRs, the olfactory receptors, are also members of the rhodopsin-like GPCR family. These receptors function by transducing odorant signals. Numerous distinct olfactory receptors are required to distinguish different odors. Each olfactory sensory neuron expresses only one type of olfactory receptor, and distinct spatial zones of neurons expressing distinct receptors are found in nasal passages. For example, the RA1c receptor which was isolated from a rat brain library, has been shown to be limited in expression to very distinct regions of the brain and a defined zone of the olfactory epithelium (Raming, K. et al. (1998) Receptors Channels 6:141-151). However, the expression of olfactory-like receptors is not confined to olfactory tissues. For example, three rat genes encoding olfactory-like receptors having typical GPCR characteristics showed expression patterns not only in taste and olfactory tissue, but also in male reproductive tissue (Thomas, M.B. et al. (1996) Gene 178:1-5).

Members of the secretin-like GPCR subfamily have as their ligands peptide hormones such as secretin, calcitonin, glucagon, growth hormone-releasing hormone, parathyroid hormone, and vasoactive intestinal peptide. For example, the secretin receptor responds to secretin, a peptide hormone that stimulates the secretion of enzymes and ions in the pancreas and small intestine (Watson, supra, pp. 278-283). Secretin receptors are about 450 amino acids in length and are found in the plasma membrane of gastrointestinal cells. Binding of secretin to its receptor stimulates the production of cAMP.

Examples of secretin-like GPCRs implicated in inflammation and the immune response include the EGF module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor proteins. These GPCRs are members of the recently characterized EGF-TM7 receptors subfamily. These seven transmembrane hormone receptors exist as heterodimers in vivo and contain between three and seven potential calcium-binding EGF-like motifs. CD97 is predominantly expressed in leukocytes and is markedly upregulated on activated B and T cells (McKnight, A.J. and S. Gordon (1998) J. Leukoc. Biol. 63:271-280).

The third GPCR subfamily is the metabotropic glutamate receptor family. Glutamate is the

major excitatory neurotransmitter in the central nervous system. The metabotropic glutamate receptors modulate the activity of intracellular effectors, and are involved in long-term potentiation (Watson, supra, p.130). The Ca^{2+}-sensing receptor, which senses changes in the extracellular concentration of calcium ions, has a large extracellular domain including clusters of acidic amino acids which may be involved in calcium binding. The metabotropic glutamate receptor family also includes pheromone receptors, the $GABA_B$ receptors, and the taste receptors.

Other subfamilies of GPCRs include two groups of chemoreceptor genes found in the nematodes Caenorhabditis elegans and Caenorhabditis briggsae, which are distantly related to the mammalian olfactory receptor genes. The yeast pheromone receptors STE2 and STE3, involved in the response to mating factors on the cell membrane, have their own seven-transmembrane signature, as do the cAMP receptors from the slime mold Dictyostelium discoideum, which are thought to regulate the aggregation of individual cells and control the expression of numerous developmentally-regulated genes.

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, supra). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) Nature 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: lutenizing hormone (precocious puberty); vasopressin V_2 (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism, hypocalcuria, hypercalcemia); parathyroid hormone (short limbed dwarfism); β_3-adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropin (glucocorticoid deficiency) (Wilson, S. et al. (1998) Br. J. Pharmocol. 125:1387-1392; Stadel, J.M. et al. (1997) Trends Pharmacol. Sci. 18:430-437). GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure, and several cardiovascular disorders (Horn, F. and G. Vriend (1998) J. Mol. Med. 76:464-468).

In addition, within the past 20 years several hundred new drugs have been recognized that are directed towards activating or inhibiting GPCRs. The therapeutic targets of these drugs span a wide range of diseases and disorders, including cardiovascular, gastrointestinal, and central nervous system disorders as well as cancer, osteoporosis and endometriosis (Wilson, supra; Stadel, supra). For example, the dopamine agonist L-dopa is used to treat Parkinson's disease, while a dopamine antagonist is used to treat schizophrenia and the early stages of Huntington's disease. Agonists and antagonists of adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and

tachycardia; serotonin 5HT1D antagonists are used against migraine; and histamine H1 antagonists are used against allergic and anaphylactic reactions, hay fever, itching, and motion sickness (Horn, supra).

Recent research suggests potential future therapeutic uses for GPCRs in the treatment of metabolic disorders including diabetes, obesity, and osteoporosis. For example, mutant V2 vasopressin receptors causing nephrogenic diabetes could be functionally rescued in vitro by co-expression of a C-terminal V2 receptor peptide spanning the region containing the mutations. This result suggests a possible novel strategy for disease treatment (Schöneberg, T. et al. (1996) EMBO J. 15:1283-1291). Mutations in melanocortin-4 receptor (MC4R) are implicated in human weight regulation and obesity. As with the vasopressin V2 receptor mutants, these MC4R mutants are defective in trafficking to the plasma membrane (Ho, G. and R.G. MacKenzie (1999) J. Biol. Chem. 274:35816-35822), and thus might be treated with a similar strategy. The type 1 receptor for parathyroid hormone (PTH) is a GPCR that mediates the PTH-dependent regulation of calcium homeostasis in the bloodstream. Study of PTH/receptor interactions may enable the development of novel PTH receptor ligands for the treatment of osteoporosis (Mannstadt, M. et al. (1999) Am. J. Physiol. 277:F665-F675).

The chemokine receptor group of GPCRs have potential therapeutic utility in inflammation and infectious disease. (For review, see Locati, M. and P.M. Murphy (1999) Annu. Rev. Med. 50:425-440.) Chemokines are small polypeptides that act as intracellular signals in the regulation of leukocyte trafficking, hematopoiesis, and angiogenesis. Targeted disruption of various chemokine receptors in mice indicates that these receptors play roles in pathologic inflammation and in autoimmune disorders such as multiple sclerosis. Chemokine receptors are also exploited by infectious agents, including herpesviruses and the human immunodeficiency virus (HIV-1) to facilitate infection. A truncated version of chemokine receptor CCR5, which acts as a coreceptor for infection of T-cells by HIV-1, results in resistance to AIDS, suggesting that CCR5 antagonists could be useful in preventing the development of AIDS.

The discovery of new G-protein coupled receptors, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, G-protein coupled receptors, referred to

collectively as "GCREC" and individually as "GCREC-1," "GCREC-2," "GCREC-3," "GCREC-4," "GCREC-5," "GCREC-6," "GCREC-7," "GCREC-8," "GCREC-9," "GCREC-10," "GCREC-11," "GCREC-12," "GCREC-13," "GCREC-14," "GCREC-15," "GCREC-16," "GCREC-17," "GCREC-18," and "GCREC-19." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-19.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-19. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:20-38.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least

90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous

nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional GCREC,

comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the

polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"GCREC" refers to the amino acid sequences of substantially purified GCREC obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GCREC. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

An "allelic variant" is an alternative form of the gene encoding GCREC. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding GCREC include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GCREC or a polypeptide with at least one functional characteristic of GCREC. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GCREC, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GCREC. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GCREC. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GCREC is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid

sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GCREC. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind GCREC polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic GCREC, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding GCREC or fragments of GCREC may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala

His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of GCREC or the polynucleotide encoding GCREC which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A

fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:20-38 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:20-38, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:20-38 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:20-38 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:20-38 and the region of SEQ ID NO:20-38 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-19 is encoded by a fragment of SEQ ID NO:20-38. A fragment of SEQ ID NO:1-19 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-19. For example, a fragment of SEQ ID NO:1-19 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-19. The precise length of a fragment of SEQ ID NO:1-19 and the region of SEQ ID NO:1-19 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of

molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Reward for match: 1
Penalty for mismatch: -2
Open Gap: 5 and Extension Gap: 2 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 11
Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for

chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of GCREC which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of GCREC which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of GCREC. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of GCREC.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of

amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an GCREC may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of GCREC.

"Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of

Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and

other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing GCREC, nucleic acids encoding GCREC, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of

replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of

the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human G-protein coupled receptors (GCREC), the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of each polypeptide of the invention, and

these properties establish that the claimed polypeptides are G-protein coupled receptors. For example, SEQ ID NO:1 is 40% identical to Meleagris gallopavo G protein-coupled P2Y nucleotide receptor (GenBank ID g2707256) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.0e-62, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a rhodopsin family 7 transmembrane receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and BLAST analyses provide further corroborative evidence that SEQ ID NO:1 is G-protein coupled receptor. SEQ ID NO:2 was analyzed and annotated in a similar manner. These analyses indicate that SEQ ID NO:2 is a pheromone receptor (Dulac, C. and R. Axel (1995) Cell 83:195-206).

As a further example, SEQ ID NO:6 is 29% identical to human C-C chemokine receptor type 1 (GenBank ID g179985) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.6e-15, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a 7 transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:6 is a chemokine receptor.

As a further example, SEQ ID NO:9 is 95% identical to rat calcium-independent alpha-latrotoxin receptor (GenBank ID g3882981) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains a 7-transmembrane receptor (secretin family) domain and a latrophilin/CL-1-like GPS domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:9 is a latrophilin-related G-protein coupled receptor.

As a further example, SEQ ID NO:12 is 84% identical to Mus musculus G-protein coupled receptor GPR73 (GenBank ID g7248884) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.7e-166, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a 7 transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis reveals the presence of a rhodopsin-like

GPCR superfamily signature (See Table 3). Additional data from MOTIFS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:12 is a G-protein coupled receptor.

As a further example, SEQ ID NO:15 is 80% identical to rat serotonin receptor (GenBank ID g310075) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.5e-152, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also contains a rhodopsin family receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, analyses provide further corroborative evidence that SEQ ID NO:15 is a G-protein coupled receptor.

As a further example, SEQ ID NO:16 is 71% identical to mouse olfactory receptor E3 (GenBank ID g3983382) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.9e-88, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a rhodopsin family 7-transmembrane receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is an olfactory G-protein coupled receptor.

As a further example, SEQ ID NO:17 is 83% identical to mouse olfactory G-protein coupled receptor G3 (GenBank ID g3983398) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.0e-99, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 also contains a rhodopsin family 7-transmembrane receptor domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:17 is an olfactory G-protein coupled receptor. SEQ ID NO:2-5, SEQ ID NO:7-8, SEQ ID NO:10-11, SEQ ID NO:13-14, and SEQ ID NO:18-19 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-19 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments

of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:20-38 or that distinguish between SEQ ID NO:20-38 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7075196H1 is the identification number of an Incyte cDNA sequence, and BRAUTDR04 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71906055V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g900324) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses GCREC variants. A preferred GCREC variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GCREC amino acid sequence, and which contains at least one functional or structural characteristic of GCREC.

The invention also encompasses polynucleotides which encode GCREC. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:20-38, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:20-38, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at

least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:20-38 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:20-38. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GCREC.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GCREC, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GCREC or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GCREC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GCREC and GCREC derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GCREC or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:20-38 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in

"Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding GCREC may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of

about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GCREC may be cloned in recombinant DNA molecules that direct expression of GCREC, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express GCREC.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GCREC-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of GCREC, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired

properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GCREC may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of GCREC, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active GCREC, the nucleotide sequences encoding GCREC or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GCREC. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GCREC. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GCREC and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-

frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GCREC and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding GCREC. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GCREC. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GCREC can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GCREC into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of

transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of GCREC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCREC may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of GCREC. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of GCREC. Transcription of sequences encoding GCREC may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GCREC may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses GCREC in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of GCREC in cell lines is preferred. For example, sequences encoding GCREC can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GCREC is inserted within a marker gene sequence, transformed cells containing sequences encoding GCREC can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GCREC under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GCREC and that express GCREC may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or

chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GCREC using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GCREC is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GCREC include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GCREC, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GCREC may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GCREC may be designed to contain signal sequences which direct secretion of GCREC through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for

post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GCREC may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric GCREC protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GCREC activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GCREC encoding sequence and the heterologous protein sequence, so that GCREC may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled GCREC may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

GCREC of the present invention or fragments thereof may be used to screen for compounds that specifically bind to GCREC. At least one and up to a plurality of test compounds may be screened for specific binding to GCREC. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of GCREC, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which GCREC binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the

compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GCREC, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing GCREC or cell membrane fractions which contain GCREC are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GCREC or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with GCREC, either in solution or affixed to a solid support, and detecting the binding of GCREC to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

GCREC of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of GCREC. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for GCREC activity, wherein GCREC is combined with at least one test compound, and the activity of GCREC in the presence of a test compound is compared with the activity of GCREC in the absence of the test compound. A change in the activity of GCREC in the presence of the test compound is indicative of a compound that modulates the activity of GCREC. Alternatively, a test compound is combined with an in vitro or cell-free system comprising GCREC under conditions suitable for GCREC activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of GCREC may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding GCREC or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-

specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding GCREC may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding GCREC can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding GCREC is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress GCREC, e.g., by secreting GCREC in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GCREC and G-protein coupled receptors. In addition, the expression of GCREC is closely associated with brain tissue, fetal brain tissue, colon polyps, diseased colon tissue, colon tumor tissue, diseased gallbladder tissue, heart tissue, diseased breast tissue, interleukin-5 stimulated eosinophils, tumor tissue, and reproductive tissues. Therefore, GCREC appears to play a role in cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections. In the treatment of disorders associated with increased GCREC expression or activity, it is desirable to decrease the expression or activity of GCREC. In the treatment of disorders associated with decreased GCREC expression or activity, it is desirable to increase the expression or activity of GCREC.

Therefore, in one embodiment, GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed

connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma,

dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus.

In another embodiment, a vector capable of expressing GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified GCREC in

conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of GCREC may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those listed above.

In a further embodiment, an antagonist of GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections described above. In one aspect, an antibody which specifically binds GCREC may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express GCREC.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of GCREC may be produced using methods which are generally known in the art. In particular, purified GCREC may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GCREC. Antibodies to GCREC may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GCREC or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic

polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GCREC have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of GCREC amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to GCREC may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GCREC-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for GCREC may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either

polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between GCREC and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering GCREC epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GCREC. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of GCREC-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GCREC epitopes, represents the average affinity, or avidity, of the antibodies for GCREC. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GCREC epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the GCREC-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of GCREC, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GCREC-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding GCREC, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding GCREC. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding GCREC. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding GCREC may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in GCREC expression or regulation causes disease, the expression of GCREC from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in GCREC are treated by constructing mammalian expression vectors encoding GCREC and introducing these vectors by mechanical means into GCREC-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii)

ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of GCREC include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). GCREC may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding GCREC from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to GCREC expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GCREC under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and

A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding GCREC to target cells which have one or more genetic abnormalities with respect to the expression of GCREC. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing GCREC to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned

herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding GCREC to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for GCREC into the alphavirus genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,

engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GCREC.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding GCREC. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding GCREC. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased GCREC expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding GCREC may be therapeutically useful, and in the treatment of disorders

associated with decreased GCREC expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GCREC may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding GCREC is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding GCREC are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding GCREC. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of

such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of GCREC, antibodies to GCREC, and mimetics, agonists, antagonists, or inhibitors of GCREC.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising GCREC or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GCREC or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and

routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GCREC or fragments thereof, antibodies of GCREC, and agonists, antagonists or inhibitors of GCREC, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GCREC may be used for the diagnosis of disorders characterized by expression of GCREC, or in assays to monitor patients being treated with GCREC or agonists, antagonists, or inhibitors of GCREC. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GCREC include methods which utilize the antibody and a label to detect GCREC in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in

the art and may be used.

A variety of protocols for measuring GCREC, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GCREC expression. Normal or standard values for GCREC expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding GCREC may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of GCREC may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GCREC, and to monitor regulation of GCREC levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GCREC or closely related molecules may be used to identify nucleic acid sequences which encode GCREC. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding GCREC, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the GCREC encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:20-38 or from genomic sequences including promoters, enhancers, and introns of the GCREC gene.

Means for producing specific hybridization probes for DNAs encoding GCREC include the cloning of polynucleotide sequences encoding GCREC or GCREC derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding GCREC may be used for the diagnosis of disorders

associated with expression of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart

disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha $_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus. The polynucleotide sequences encoding GCREC may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like

assays; and in microarrays utilizing fluids or tissues from patients to detect altered GCREC expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GCREC may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GCREC may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding GCREC in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of GCREC, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GCREC, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GCREC may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide

encoding GCREC, or a fragment of a polynucleotide complementary to the polynucleotide encoding GCREC, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of GCREC include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor

progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, GCREC, fragments of GCREC, or antibodies specific for GCREC may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is

not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the

polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for GCREC to quantify the levels of GCREC expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci.

USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding GCREC may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GCREC on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of

the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, GCREC, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GCREC and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with GCREC, or fragments thereof, and washed. Bound GCREC is then detected by methods well known in the art. Purified GCREC can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GCREC specifically compete with a test compound for binding GCREC. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GCREC.

In additional embodiments, the nucleotide sequences which encode GCREC may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/221,478, U.S. Ser. No. 60/223,268, U.S. Ser. No. 60/231,121, U.S. Ser. No. 60/232,691, U.S. Ser. No. 60/235,146, U.S. Ser. No. 60/227,054, and U.S. Ser. No. 60/232,243, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database

(Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on

GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:20-38. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative G-protein coupled receptors were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been

annotated as G-protein coupled receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public

databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of GCREC Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:20-38 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:20-38 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding GCREC are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and

disease-specific expression of cDNA encoding GCREC. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extensi n of GCREC Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%)

agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:20-38 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and

compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified

using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is

incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each

spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the GCREC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCREC. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GCREC. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the GCREC-encoding transcript.

XII. Expression of GCREC

Expression and purification of GCREC is achieved using bacterial or virus-based expression systems. For expression of GCREC in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express GCREC upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GCREC in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding GCREC by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, GCREC is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on

immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GCREC at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified GCREC obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

GCREC function is assessed by expressing the sequences encoding GCREC at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GCREC on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GCREC and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake

Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding GCREC and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of GCREC Specific Antibodies

GCREC substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the GCREC amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-GCREC activity by, for example, binding the peptide or GCREC to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring GCREC Using Specific Antibodies

Naturally occurring or recombinant GCREC is substantially purified by immunoaffinity chromatography using antibodies specific for GCREC. An immunoaffinity column is constructed by covalently coupling anti-GCREC antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing GCREC are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GCREC (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/GCREC binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and GCREC is collected.

XVI. Identification of Molecules Which Interact with GCREC

GCREC, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GCREC, washed,

and any wells with labeled GCREC complex are assayed. Data obtained using different concentrations of GCREC are used to calculate values for the number, affinity, and association of GCREC with the candidate molecules.

Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

GCREC may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of GCREC Activity

An assay for GCREC activity measures the expression of GCREC on the cell surface. cDNA encoding GCREC is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using GCREC-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of GCREC expressed on the cell surface.

In the alternative, an assay for GCREC activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding GCREC is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [^{3}H]thymidine, a radioactive DNA precursor molecule. Varying amounts of GCREC ligand are then added to the cultured cells. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GCREC ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of GCREC producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for GCREC activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) J. Biol. Chem. 273:4990-4996). A plasmid encoding full length

GCREC is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing $1x10^5$ cells/well and incubated with inositol-free media and [^{3}H]myoinositol, 2 μCi/well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

XVIII. Identification of GCREC Ligands

GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca^{2+}. These may be measured directly using standard methods well known in the art, or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) Trends Pharmacol. Sci. 17:235-237). Assay technologies are available for both of these second messenger systems to allow high throughput readout in multi-well plate format, such as the adenylyl cyclase activation FlashPlate Assay (NEN Life Sciences Products), or fluorescent Ca^{2+} indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, GCREC may be coexpressed with the G-proteins $G_{\alpha 15/16}$ which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel the signal transduction of the GCREC through a pathway involving phospholipase C and Ca^{2+}

mobilization. Alternatively, GCREC may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCREC activation screening. These yeast systems substitute a human GPCR and Ga protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways are also modified so that the normal yeast response to the signal is converted to positive growth on selective media or to reporter gene expression (Broach, J.R. and J. Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7474806	1	7474806CD1	20	7474806CB1
7474840	2	7474840CD1	21	7474840CB1
7475092	3	7475092CD1	22	7475092CB1
7341260	4	7341260CD1	23	7341260CB1
7473911	5	7473911CD1	24	7473911CB1
7474767	6	7474767CD1	25	7474767CB1
7475815	7	7475815CD1	26	7475815CB1
60263275	8	60263275CD1	27	60263275CB1
60203310	9	60203310CD1	28	60203310CB1
7477349	10	7477349CD1	29	7477349CB1
55002225	11	55002225CD1	30	55002225CB1
7475686	12	7475686CD1	31	7475686CB1
7482007	13	7482007CD1	32	7482007CB1
6769042	14	6769042CD1	33	6769042CB1
7476053	15	7476053CD1	34	7476053CB1
7480410	16	7480410CD1	35	7480410CB1
55036418	17	55036418CD1	36	55036418CB1
7481701	18	7481701CD1	37	7481701CB1
7481774	19	7481774CD1	38	7481774CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	7474806CD1	g2707256	4.00E-62	[fl][Meleagris gallopavo] G protein coupled P2Y nucleotide receptor
2	7474840CD1	g11967419	4.00E-65	[fl][Mus musculus] vomeronasal (pheromone) receptor V1RC3
3	7475092CD1	g2992628	2.00E-86	[fl][Homo sapiens] putative seven pass transmembrane protein
4	7341260CD1	g3805932	1.00E-10	[fl][Homo sapiens] putative G-Protein coupled receptor, EDG6 Graler, M.H. et al. (1998) Genomics 53:164-169
5	7473911CD1	g1055254	2.00E-18	[fl][Rattus norvegicus] pheromone receptor VN6 Dulac, C. and Axel, R. (1995) Cell 83: 195-206
6	7474767CD1	g179985	2.00E-14	[fl][Homo sapiens] C-C chemokine receptor type 1 Neote, K. et al. (1993) Cell 72:415-425
7	7475815CD1	g1055254	1.00E-56	[fl][Rattus norvegicus] pheromone receptor VN6
8	60263275CD1	g13183149	0	[fl][Homo sapiens] (AF239764) EGF-like module-containing mucin-like receptor EMR3
9	60203310CD1	g3882981	0	[fl][Rattus norvegicus] calcium-independent alpha-latrotoxin receptor
10	7477349CD1	g6979162	3.00E-11	[fl][Rattus norvegicus] macrophage inflammatory protein-1 alpha receptor
11	55002225CD1	g14164383	0	[fl][Homo sapiens] (AB060151) G protein-coupled receptor
12	7475686CD1	g7248884	1.00E-180	[fl][Mus musculus] G-protein coupled receptor GPR73 Parker, R. et al. (2000) Biochim. Biophys. Acta 1491:369-375
13	7482007CD1	g5525078	1.00E-104	[fl][Rattus norvegicus] seven transmembrane receptor Abe, J. et al. (1999) J. Biol. Chem. 274:19957-19964
14	6769042CD1	g4164061	1.70E-59	[Bos taurus] latrophilin 3 splice variant abbg FEBS Lett. (1999) 443:348-352

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
15	7476053CD1	g310075	1.00E-165	[fl][Rattus norvegicus] serotonin receptor Erlander, M.G. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:3452-3456
16	7480410CD1	g3983382	1.90E-88	[Mus musculus] olfactory receptor E3
17	55036418CD1	g3983398	5.00E-99	[Mus musculus] olfactory receptor G3 Krautwurst, D. et al. (1998) Cell 95:917-926
18	7481701CD1	g12007416	2.00E-67	[fl][Mus musculus] m51 olfactory receptor
19	7481774CD1	g5901478	3.10E-106	[Marmota marmota] olfactory receptor

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7474806CD1	339	S111 S15 S189 S260 S336 S46 S84 T163 T204 T255 T36 T80	N35 N69	Transmembrane domains: Y89-T109, F130-F149	HMMER
					7 transmembrane receptor (rhodopsin family): L68-Y324	HMMER-PFAM
					G-protein coupled receptor BL00237: W116-P155, F224-Y235, T255-T281, S316-I332	BLIMPS-BLOCKS
					Rhodopsin-like GPCR superfamily PR00237: Y53-W77, A86-I107, F130-V152, H166-A187, G216-V239, S260-F284, Y306-I332	BLIMPS-PRINTS
					G-protein coupled receptor: DM00013 P41231|27-322: F45-S327 P51582|29-322: F45-S327 I55450|20-317: F45-R328 P48042|45-340: Y53-S327	BLAST-DOMO
					G-protein coupled receptor: PD000009: Y89-F188	BLAST-PRODOM
2	7474840CD1	335	S237 S28 S318 T22 T86	N168 N184	Pheromone receptor: PD009900: I59-M329	BLAST-PRODOM
3	7475092CD1	428	S257 S286 S295 S342 S356 S365 S412 S79 T156	N180 N207 N284	Transmembrane domains: H166-L186, M240-Y256, Y299-L319	HMMER
					Putative seven pass transmembrane protein: PD138976: R3-L360	BLAST-PRODOM
4	7341260CD1	330	S123 S185 S323 S5 S78 T219	N154 N4 N76 N93	Transmembrane Domain: L55-L74	HMMER
					7 transmembrane receptor (rhodopsin family): A31-A250	HMMER_PFAM
					Rhodopsin-like GPCR superfamily signature PR00237B:A50-L71 PR00237C:P92-V114 PR00237D:S123-L144 PR00237F:Q222-L246 PR00237G:L262-T288	BLIMPS_PRINTS
5	7473911CD1	676	S105 S213 S215 S250 S330 S353 S492 S569 S591 S606 S646 S647 S79 T217 T296 T301 T394 T620	N184	Transmembrane Domain: I118-A136	HMMER
					PHEROMONE RECEPTOR PD009900: S87-V167, K42-Y90	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	7474767CD1	372	S214 S289 S351 T145 T223 T299	N11 N318 N353	Signal Peptide: M1-T47	SPSCAN
					Transmembrane Domains: Y32-S52, I74-P100, I184-Y208	HMMER
					7 transmembrane receptor (rhodopsin family): A43-Y285	HMMER_PFAM
					G-protein coupled receptors proteins BL00237:M94-P133, Y221-M247, N277-R293	BLIMPS_BLOCKS
					Rhodopsin-like GPCR superfamily signature PR00237: V28-S52, S62-V83, E108-V130, R144-W165, I184-V207, T226-Y250, S267-R293	BLIMPS_PRINTS
					G-protein coupled receptors signature: E105-T155	PROFILESCAN
					G-PROTEIN COUPLED RECEPTORS DM00013 P32246|28-316:A22-R293 P51677|28-316:F27-R293 P25930|33-319:F27-L300 P34981|19-336:V28-I206, S222-R293	BLAST_DOMO
7	7475815CD1	271	S115 S142 S188 S203 S79 T269		Transmembrane Domain: I216-A234	HMMER
					PHEROMONE RECEPTOR PD009900: K42-S170, I112-K271	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	60263275CD1	611	S257 S308 S494 S578 S580 S588 S598 S63 S81 T21 T211 T258 T307 T406 T75 Y585	N104 N148 N161 N209 N238 N286 N293 N345 N409 N414	Signal Peptide: M1-Q22	HMMER
					Transmembrane Domains: V321-L339, W424-T451, M464-I488, Y540-L560	HMMER
					7 transmembrane receptor (Secretin family): D312-V564	HMMER_PFAM
					Latrophilin/CL-1-like GPS domain: K259-Q309	HMMER_PFAM
					EGF-like domain: C30-C76	HMMER_PFAM
					G-protein coupled receptor BL00649: C378-L403, G430-R454, W465-S494	BLIMPS_BLOCKS
					Secretin-like GPCR superfamily signature PR00249: V317-K341, I380-L403, K423-L448 W465-K490, A539-L560	BLIMPS_PRINTS
					Type I EGF signature PR00009: K25-F40, E48-Y59	BLIMPS_PRINTS
					G-protein coupled receptors family 2 signatures: V531-E574	PROFILESCAN
					RECEPTOR G PROTEIN COUPLED PD000752: L315-F572	BLAST_PRODOM
					RECEPTOR G PROTEIN COUPLED EGF LIKE PD005428:N176-M304	BLAST_PRODOM
					HORMONE; EMR1; LEUCOCYTE; ANTIGEN; DM05221 A57172|465-886:S183-K596 I37225|347-738:K259-G592 P48960|347-738:K259-G592	BLAST_DOMO
					G-PROTEIN COUPLED RECEPTORS FAMILY 2 DM00378|I49149|56-425:V314-R573, G9-V56	BLAST_DOMO
					Aspartic acid and asparagine hydroxylation site: C44-C55	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	60203310CD1	1469	S166 S186 S1214 S187 S261 T1253 S325 S351 S1295 S417 S423 T1398 S460 S591 T1453 S630 S761 T1008 S791 S824 S1144 S88 S907 S1154 S978 T113 T1181 T181 T199 S1295 T294 T303 T1346 T443 T559 S1458 T658 T672 T731 T803 T842 Y121	N464 N549 N759 N772 N817 N843 N93 N932 N1098 N1179 N1196 N1251	Signal peptide: M1-S21	HMMER
					Transmembrane domains: V879-F897, V942-L960, F1020-G1040, F1071-W1090	HMMER
					7 transmembrane receptor (secretin family): D874-V1130	HMMER-PFAM
					Latrophilin/CL-1-like GPS domain: F814-V866	HMMER-PFAM
					G-protein coupled receptor BL00649: C940-L965, G987-V1011, W1022-A1051, S1113-L1138, C498-T525, G884-I929	BLIMPS-BLOCKS
					Secretin-like GPCR superfamily PR00249: V879-R903, V942-L965, R980-S1005, W1022-V1047, A1105-L1126	BLIMPS-PRINTS
					EMR1 hormone leucocyte antigen: DM05221 I37225|347-738: L769-S1160 P48960|347-738: L769-S1160 A57172|465-886: N730-T1162	BLAST-DOMO
					G-protein coupled receptors family 2: DM00378|I49149|56-425: G836-R1131	BLAST-DOMO
					Latrophilin-related receptor: PD024331: K386-E628 PD041747: L1207-L1469	BLAST-PRODOM
					Myocilin/Olfactomedin response protein: PD006897: Y177-V385	BLAST-PRODOM
					EGF-like G-protein coupled receptor: PD005428: R629-V866	BLAST-PRODOM
10	7477349CD1	469	S122 S199 S25 S312 S447 S85 T110 T195 T207 T282 T315 T39 T427 T431 T465	N139	Transmembrane domains: P166-L188, V215-V233	HMMER
					7 transmembrane receptor (rhodopsin family): A251-Y421	HMMER-PFAM
					G-protein coupled receptor motif: A251-L267	MOTIFS
					G-protein coupled receptors signature: I243-A293	ProfileScan
					G-protein coupled receptor BL00237: R231-P270, Y329-N340, P357-V383, H413-R429	BLIMPS-BLOCKS
					Rhodopsin-like GPCR superfamily: I165-A189, S199-A220, E245-L267, R281-W302, K321-I344, S362-Y386,L403-R429	BLIMPS-PRINTS
					G-protein coupled receptor: DM00013	BLAST-DOMO
					I49339|28-316: A162-V435	BLAST-DOMO
					P51675|28-317: A162-V435	

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	55002225CD1	335	S190 S30 S62 T11 T110 T150 T189 T320 T330 T66	N10 N17	G-PROTEIN COUPLED RECEPTORS DM00013|P32745|38-329:V32-S310	BLAST_DOMO
					RECEPTOR COUPLED G-PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE PROTEIN FAMILY PD000009:R61-Y171	BLAST_PRODOM
					G-protein coupled receptor BL00237:S298-Q314, W99-P138, F206-Y217, Q244-I270	BLIMPS_BLOCKS
					Rhodopsin-like GPCR superfamily signature PR00237:I36-I60, P68-L89, D113-L135, K149-V170, T198-L221, L249-V273, Y288-Q314	BLIMPS_PRINTS
					Transmembrane domain transmem_domain:T34-F57, V194-I218, M252-L272	HMMER
					7 transmembrane receptor (rhodopsin family) 7tm_1:G51-Y306	HMMER_PFAM
					G_Protein_Receptor A119-L135	MOTIFS
					G-protein coupled receptors signature g_protein_receptor.prf:T110-L157	PROFILESCAN
12	7475686CD1	630	S230 S269 S601 S607 T223 T519 T581 T586 T620 T81	N332 N584	G-PROTEIN COUPLED RECEPTORS DM00013|P49146|44-341:K299-F590	BLAST_DOMO
					RECEPTOR COUPLED G-PROTEIN TRANSMEMBRANE GLYCOPROTEIN PHOSPHORYLATION LIPOPROTEIN PALMITATE PROTEIN FAMILY PD000009:K328-F437	BLAST_PRODOM
					G-protein coupled receptor BL00237:W367-P406, F478-Y489, R514-F540	BLIMPS_BLOCKS
					Rhodopsin-like GPCR superfamily signature PR00237:R381-I403, T415-Y436, F470-S493, T519-V543, F561-M587, V301-T325, T334-F355.	BLIMPS_PRINTS
					Neuropeptide Y receptor PR01012:R326-I338, R381-A396	BLIMPS_PRINTS
					Transmembrane domain transmem_domain: I302-F320	HMMER
					7 transmembrane receptor (rhodopsin family) 7tm_1:G316-N575	HMMER_PFAM
					G_Protein_Receptor V387-I403	MOTIFS
					G-protein coupled receptors signature g_protein_receptor.prf:N378-I428	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	7482007CD1	695	S185 S252 S354 S392 S395 S591 S669 T265 T397 T45 T584 T624 T653	N169 N177 N209 N229 N250 N257 N263 N286 N309 N340 N379 N61 N679 N686	HORMONE; EMR1; LEUCOCYTE; ANTIGEN; DM05221|I37225|347-738:C349-S688	BLAST_DOMO
					RECEPTOR TRANSMEMBRANE G-PROTEIN COUPLED GLYCOPROTEIN PRECURSOR SIGNAL TYPE POLYPEPTIDE ALTERNATIVE PD000752:N372-R660	BLAST_PRODOM
					G-protein coupled receptors family 2 proteins. BL00649:C473-I498	BLIMPS_BLOCKS
					Secretin-like GPCR superfamily signature PR00249:Y403-W427, A475-I498	BLIMPS_PRINTS
					Transmembrane domain transmem_domain:I407-T433, M514-T537, L487-F506, L559-V579	HMMER
					7 transmembrane receptor (Secretin family) 7tm_2:D398-I659	HMMER_PFAM
14	6769042CD1	633	S108 S262 S275 S381 S589 S612 S616 S70 T251 T355 T517	N153 N235 N260 N263 N292 N366 N41 N604 N61 N78	HORMONE; EMR1; LEUCOCYTE; ANTIGEN; DM05221|A57172|465-886:G279-S593	BLAST_DOMO
					RECEPTOR TRANSMEMBRANE G-PROTEIN COUPLED GLYCOPROTEIN PRECURSOR SIGNAL TYPE POLYPEPTIDE ALTERNATIVE PD000752:H321-K579	BLAST_PRODOM
					G-protein coupled receptors family 2 proteins. BL00649:S555-I580, C391-L416	BLIMPS_BLOCKS
					cAMP-type GPCR signature PR00247:Y361-R383, A395-I421, Y433-S451	BLIMPS_PRINTS
					Secretin-like GPCR superfamily signature PR00249:S327-S351, V393-L416, H431-S456, W473-S498, T517-L537, Q547-L568	BLIMPS_PRINTS
					Transmembrane domain transmem_domain: C333-L350, I472-T494	HMMER
					7 transmembrane receptor (Secretin family) 7tm_2:Q322-V572	HMMER_PFAM
					Latrophilin/CL-1-like GPS domain GPS:V265-L315	HMMER_PFAM
					G-protein coupled receptors family 2 signatures g_protein_recep_f2_2.prf: A474- S593	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databas s
15	7476053CD1	370	S217 S27 S287 T116 T163		Signal peptide: M298-T317	HMMER
					Signal_cleavage: M1-A65	SPSCAN
					Transmembrane domain: F48-L68, M298-L316	HMMER
					7 transmembrane receptor (rhodopsin family) 7tm_1: W69-Y351	HMMER_PFAM
					G-protein coupled receptor BL00237: R120-H159, Q291-T317, N343-N359	BLIMPS_BLOCKS
					Rhodopsin-like GPCR superfamily signature PR00237: L54-P78,P87-P108, D134-I156, A170-L191, A214-Y237, A296-I320, K333-N359	BLIMPS_PRINTS
					5-hydroxytryptamine 5B receptor PR00519: A3-P19, E20-P36, P36-V50, E196-R204, R246-V260, V260-V270	BLIMPS_PRINTS
					5HYDROXYTRYPTAMINE 5B RECEPTOR 5HT5B SEROTONIN GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE PD027821: M1-H84	BLAST_PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013 P31387 | 46-367: P46-T367 I48231 | 46-367: P46-T367 P47898 | 34-354: P46-T367 P20905 | 156-522: P46-S280	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	7480410CD1	324	S189 S194 S292 S314 S68	N9	Signal peptide: M1-A39	SPScan
					Transmembrane domains:G26-I50, C203-V217	HMMER
					7 transmembrane receptor (rhodopsin family): A42-Y291	HMMER-PFAM
					G-protein coupled receptors signature: F103-V147	ProfileScan
					G-protein coupled receptor motif: G111-V127	MOTIFS
					G-protein coupled receptor BL00237: T283-A299, K91-P130, L208-Y219, R236-L262	BLIMPS-BLOCKS
					Rhodopsin-like GPCR superfamily PR00237: L27-H51, M60-K81, Y105-V127, M200-L223, P25-L49, K273-A299	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M60-K81, F178-D192, F239-G254, V275-L286, S292-L306	BLIMPS-PRINTS
					G-protein coupled receptor: DM00013 P23275|17-306: H21-G307 A57069|15-304: H21-G307 P30954|29-316: L28-G307 S29709|11-299: P25-G307	BLAST-DOMO
					Olfactory G-protein coupled receptor: PD000921: F169-M247	BLAST-PRODOM
					Olfactory G-protein coupled receptor: PD149621: V248-R308	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	5503641 8CD1	315	S194 S22 S292 S68 S8 S88	N6	Signal peptide: M1-A24	SPScan
					Transmembrane domains: L26-L46, C98-M119, F201-A220	HMMER
					7 transmembrane receptor (rhodopsin family): G42-Y291	HMMER-PFAM
					G-protein coupled receptor motif: S111-I127	MOTIFS
					G-protein coupled receptors signature: L104-G147	ProfileScan
					G-protein coupled receptor BL00237: K91-P130, L208-Y219, K236-R262, T283-M299	BLIMPS-BLOCKS
					Rhodopsin-like GPCR superfamily PR00237: V27-Y51, M60-K81, F105-I127, I200-L223, K273-M299	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M60-K81, F178-D192, L239-G254, A275-L286, S292-L306	BLIMPS-PRINTS
					G-protein coupled receptor: DM00013 P23275 17-306: I18-L302 A57069 15-304: F19-K304 P34982 17-305: V27-D307 P23270 18-311: L26-L306	BLAST-DOMO
					Olfactory G-protein coupled receptor: PD000921: Y169-L246	BLAST-PRODOM
					Olfactory G-protein coupled receptor: PD149621: T247-R308	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7481701CD1	324	S292 S314 S320 S67 S87 T136 T8		Transmembrane domains: L32-N52, V243-F263, Y267-N287	HMMER
					7 transmembrane receptor (rhodopsin family): G41-F180, V225-Y291	HMMER-PFAM
					Visual pigments (opsins) retinal binding site: W271-T319	ProfileScan
					G-protein coupled receptor BL00237: R89-P128, R236-R262, S283-K299	BLIMPS-BLOCKS
					Rhodopsin-like GPCR superfamily PR00237: V26-Q50, M59-K80, F103-I125, F12-F33, T144-F167, A238-R262, I273-K299	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M59-K80, I176-D190, F239-G254, S292-I306	BLIMPS-PRINTS
					Melanocortin receptor family PR00534: L51-I63, I125-T136	BLIMPS-PRINTS
					G-protein coupled receptor: DM00013 P23267|20-309: F17-I306 S29709|11-299: V29-G307 P23270|18-311: F17-K304 P23274|18-306: V26-I302	BLAST-DOMO
					Olfactory G-protein coupled receptor: PD000921: L165-I247	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	7481774CD1	312	S16 S230 S264 S47 S65 T161 T289	N40	Transmembrane domains: I24-V46, Q98-M116, F198-L214	HMMER
					7 transmembrane receptor (rhodopsin family): G39-Y288	HMMER-PFAM
					G-protein coupled receptor motif: V108-I124	MOTIFS
					G-protein coupled receptors signature: F100-T146	ProfileScan
					G-protein coupled receptor BL00237: K88-P127, V205-Y216, Q233-Q259, T280-K296	BLIMPS-BLOCKS
					Olfactory receptor signature PR00245: M57-Q78, F175-S189, F236-G251, L272-L283, T289-L303	BLIMPS-PRINTS
					G-protein coupled receptor: DM00013 P23275 | 17-306: S16-L303 A57069 | 15-304: F15-L303 P34982 | 17-305: S16-L303 P30953 | 18-306: R18-L303	BLAST-DOMO
					Olfactory G-protein coupled receptor: PD149621: T244-R307	BLAST-PRODOM
					Olfactory G-protein coupled receptor: PD000921: C167-L243	BLAST-PRODOM

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
20	7474806CB1	1076	1-1076	7075196H1 (BRAUTDR04)	1	532
				g7248967_edit	533	1076
21	7474840CB1	1102	1007-1102, 107-160, 486-522, 642-842	g7407927_edit	1	1102
22	7475092CB1	2529	628-1475, 1-81	55049853H1	876	1612
				71906055V1	564	1209
				1351856F6 (LATRTUT02)	1757	2367
				GNN.g7023955_000031_002	1	468
				71900320V1	1988	2529
				8017335J1 (BMARTXE01)	64	614
				55049805J1	1256	1956
23	7341260CB1	1847	1-140, 1490-1847	7341260H1 (COLNDIN02)	307	991
				70811587V1	1239	1847
				6834479H1 (BRSTNON02)	974	1575
				70888652V1	228	660
				7628842H1 (GBLADIE01)	1	284
24	7473911CB1	2031	1-504, 669-1834	FL140044_00001	1	2031
25	7474767CB1	1130	1-1130	GNN.g6693326_000106_002	139	1130
				55093139J1	1	276
26	7475815CB1	1202	367-959, 1044-1202, 1-116	CpG_991027_B15_masked_fa.	1	723
				FL7475815_g8492585_000004_g3892596	372	1202
27	60263275CB1	2079	134-490, 902-981	71704087V1	85	700
				71651942V1	986	1727
				3642425T6 (LUNGNOT34)	1469	2079
				524802R6 (CARCTXT01)	391	985
				2435123H1 (BRAVUNT02)	1	255
				71651560V1	838	1544

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
28	60203310CB1	5324	1-373, 963-3661, 4111-4162	71959831V1	4179	5068
				60203311D1	3309	3622
				7638002J1 (SEMVTDE01)	2283	2900
				491493H1 (HNT2AGT01)	3833	4073
				71957696V1	4646	5324
				4028716F7 (BRAINOT23)	4057	4787
				8103587H1 (MIXDDIE02)	620	1259
				g900324	3652	4165
				8195081H2 (BRAIDIR04)	1	623
				55073390J1	1549	2391
				60200671D1	3444	3823
				8195081J2 (BRAIDIR04)	432	1033
				55094091J1	1140	2026
				55116120H1	2658	3416
29	7477349CB1	1962	1-1758	70846216V1	1486	1962
				GNN.g8140731_000028_002	265	1410
				GNN.g9309533_000030_002	1	379
				2021568F6 (CONNNOT01)	379	873
				71243436V1	1272	1931
				70844223V1	872	1416
				70845761V1	804	1415
30	55002225CB1	1558	80-1558, 1-49	72398219V1	675	1384
				72374379V1	971	1558
				72373094V1	814	1540
				55049494J1	1	817
31	7475686CB1	2304	1-1212, 1548-2304	CpG_SAE300482544.R1	1148	1481
				GNN:g6138786_000002_006.edit	1	1893
				7290466F6 (BRAIFER06)	1576	2304
32	7482007CB1	2322	1-628, 1452-1878, 724-1278	55049456J1	461	1327
				6925371H1 (PLACFER06)	249	797
				GNN:g9864547_000007_002_edit	472	2103
				55084155J2	1017	1790
				7341368F8 (COLNDIN02)	1	440
				g1507289	1879	2322

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
33	6769042CB1	2366	800-1248, 1-40, 1369-1717, 218-361, 1821-1861, 1930-2366	7157074H2 (BRAIFEJ02)	1169	1797
				72138116D1	1704	2366
				7629227H1 (GBLADIE01)	834	1341
				55147608J1	1	908
34	7476053CB1	1458	1-1087	GNN.g7630808_000013_022	1201	1414
				g3280262	1088	1458
				GBI:g9454621.edit	1	1110
35	7480410CB1	975	1-816, 894-975	55036194H1 (GPCRDPV02)	166	371
				GNN.g8979559_000011_002	1	975
36	55036418CB1	948	1-162, 193-948	55036391H1 (GPCRDPV02)	166	371
				GNN.g7239420_000072_008	1	948
37	7481701CB1	1086	1-1086	GNN.g9795014_000002_002	1	1086
38	7481774CB1	1529	1-963	55143535J1	127	918
				70822063V1	879	1529
				7361408F8 (BRAIFEE05)	401	1051
				55142634H1	1	384

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
20	7474806CB1	BRAUTDR04
22	7475092CB1	LATRTUT02
23	7341260CB1	COLNTUT03
24	7473911CB1	BRSTNOT23
27	60263275CB1	EOSITXT01
28	60203310CB1	BRAITUT01
29	7477349CB1	CONNNOT01
31	7475686CB1	BRAIFER06
32	7482007CB1	COLNDIN02
33	6769042CB1	GBLADIE01
38	7481774CB1	BRAIFEE05

Table 6

Library	Vector	Library Description
BRAIFEE05	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from brain tissue removed from Caucasian male fetus who was still born with a hypoplastic l ft heart at 23 weeks' gestation. Serologies were negative.
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRAITUT01	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from a 50-year-old Caucasian female during a frontal lobectomy. Pathology indicated recurrent grade 3 oligoastrocytoma with focal necrosis and extensive calcification. Patient history included a speech disturbance and epilepsy. The patient's brain had also been irradiated with a total dose of 5,082 cyg (Fraction 8). Family history included a brain tumor.
BRAUTDR04	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled striatum, dorsal caudate nucleus, dorsal putamen, and ventral nucleus accumbens tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydorthorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTNOT23	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 35-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated nonproliferative fibrocystic disease. Family history included type II diabetes, atherosclerotic coronary artery disease, acute myocardial infarction, hyperlipidemia, and coronary artery bypass.
COLNDIN02	pINCY	This normalized library was constructed from 4.72 million independent clones from a diseased colon and colon polyp tissue library. Starting RNA was made from pooled cDNA from two donors. cDNA was generated using mRNA isolated from diseased colon tissue removed from the cecum and descending colon of a 16-year-old Caucasian male (donor A) during partial colectomy, temporary ileostomy, and colonoscopy and from diseased colon polyp tissue removed from the cecum of a 67-year-old female (donor B). Pathology indicated innumerable (greater than 100) adenomatous polyps with low-grade dysplasia involving the entire colonic mucosa in the setting of familial polyposis coli (donor A), and a benign cecum polyp (donor B). Pathology for the associated tumor tissue (B) indicated invasive grade 3 adenocarcinoma that arose in tubulovillous adenoma forming a fungating mass in the cecum. The tumor infiltrated just through the muscularis propria. Multiple (2 of 17) regional lymph nodes were involved by metastatic adenocarcinoma. A tubulovillous adenoma and multiple (6) tubular adenomas with low-grade dysplasia were observed in the cecum and ascending colon. Donor A presented with abdominal pain and

Table 6 (cont.)

Library	Vector	Library Description
		flatulence. The patient was not taking any medications. Family history included benign colon neoplasm in the father and sibling(s); benign hypertension, cerebrovascular disease, breast cancer, uterine cancer, and type II diabetes in the grandparent(s).
COLNTUT03	pINCY	Library was constructed using RNA isolated from colon tumor tissue obtained from the sigmoid colon of a 62-year-old Caucasian male during a sigmoidectomy and permanent colostomy. Pathology indicated invasive grade 2 adenocarcinoma. One lymph node contained metastasis with extranodal extension. Patient history included hyperlipidemia, cataract disorder, and dermatitis. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, breast cancer , and prostate cancer.
CONNNOT01	pINCY	Library was constructed using RNA isolated from mesentery fat tissue obtained from a 71-year-old Caucasian male during a partial colectomy and permanent colostomy. Family history included atherosclerotic coronary artery disease, myocardial infarction, and extrinsic asthma.
EOSITXT01	pINCY	Library was constructed using RNA isolated from eosinophils stimulated with IL-5.
GBLADIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased gallbladder tissue removed from a 55-year-old Caucasian female during laparoscopic cholecystectomy. Pathology indicated chronic cholecystitis and cholelithiasis (greater than 100 stones). The patient presented with cholelithiasis, abdominal pain, and tremors. Patient history included benign hypertension, Morton's neuroma, facial hirsutism, normal delivery, and tobacco abuse in remission. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, and adenotonsillectomy. Patient medications included Inderal and Premarin. Family history included breast cancer and ALS in the mother; chronic leukemia and ARDS in the father; breast cancer in the sibling(s); and atherosclerotic coronary artery disease in the grandparent(s).
LATRTUT02	pINCY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-19,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-19.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:20-38.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

18. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim

1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the

amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of GCREC in a biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of GCREC in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of GCREC in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, or an immunogenic fragment thereof, under conditions to

elicit an antibody response;

b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:20.

65. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:21.

66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.

67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.

68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.

69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.

74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.

79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.

81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID

NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.

<110> INCYTE GENOMICS, INC.
THORNTON, Michael
PATTERSON, Chandra
LAL, Preeti
BURFORD, Neil
YUE, Henry
GANDHI, Ameena R.
ELLIOTT, Vicki S.
RAMKUMAR, Jayalaxmi
BAUGHN, Mariah R.
KALLICK, Deborah A.
WALIA, Narinder K.
HAFALIA,April J.A.
YAO, Monique G.
LU, Yan
TRIBOULEY, Catherine M.
POLICKY, Jennifer L.
KEARNEY, Liam
GRAUL, Richard
WARREN, Bridget
LEE, Ernestine A.
DING, Li

<120> G-PROTEIN COUPLED RECEPTORS

<130> PI-0176 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/221,478; 60/223,268; 60/227,054; 60/231,121; 60/232,243;
60/232,691; 60/235,146
151> 2000-07-27; 2000-08-03; 2000-08-21; 2000-09-08; 2000-09-13;
2000-09-15; 2000-09-22

<160> 38
<170> PERL Program

<210> 1
<211> 339
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474806CD1

<400> 1

Met	Leu	Ser	Ile	Leu	Leu	Pro	Ser	Arg	Gly	Ser	Arg	Ser	Gly	Ser	
1				5				10						15	
Arg	Arg	Gly	Ala	Leu	Leu	Leu	Glu	Gly	Ala	Ser	Arg	Asp	Met	Glu	
				20				25						30	
Lys	Val	Asp	Met	Asn	Thr	Ser	Gln	Glu	Gln	Gly	Leu	Cys	Gln	Phe	
				35				40						45	
Ser	Glu	Lys	Tyr	Lys	Gln	Val	Tyr	Leu	Ser	Leu	Ala	Tyr	Ser	Ile	
				50				55						60	
Ile	Phe	Ile	Leu	Gly	Leu	Pro	Leu	Asn	Gly	Thr	Val	Leu	Trp	His	
				65				70						75	
Ser	Trp	Gly	Gln	Thr	Lys	Arg	Trp	Ser	Cys	Ala	Thr	Thr	Tyr	Leu	
				80				85						90	
Val	Asn	Leu	Met	Val	Ala	Asp	Leu	Leu	Tyr	Val	Leu	Leu	Pro	Phe	
				95				100						105	
Leu	Ile	Ile	Thr	Tyr	Ser	Leu	Asp	Asp	Arg	Trp	Pro	Phe	Gly	Glu	

```
                110                 115                 120
Leu Leu Cys Lys Leu Val His Phe Leu Phe Tyr Ile Asn Leu Tyr
                125                 130                 135
Gly Ser Ile Leu Leu Leu Thr Cys Ile Ser Val His Gln Phe Leu
                140                 145                 150
Gly Val Trp His Pro Leu Cys Ser Leu Pro Tyr Arg Thr Arg Arg
                155                 160                 165
His Ala Trp Leu Gly Thr Ser Thr Thr Trp Ala Leu Val Val Leu
                170                 175                 180
Gln Leu Leu Pro Thr Leu Ala Phe Ser His Thr Asp Tyr Ile Asn
                185                 190                 195
Gly Gln Met Ile Trp Tyr Asp Met Thr Ser Gln Glu Asn Phe Asp
                200                 205                 210
Arg Leu Phe Ala Tyr Gly Ile Val Leu Thr Leu Ser Gly Phe Leu
                215                 220                 225
Ser Pro Ser Leu Val Ile Leu Val Cys Tyr Ser Leu Met Val Arg
                230                 235                 240
Ser Leu Ile Lys Pro Glu Glu Asn Leu Met Arg Thr Gly Asn Thr
                245                 250                 255
Ala Arg Ala Arg Ser Ile Arg Thr Ile Leu Leu Val Cys Gly Leu
                260                 265                 270
Phe Thr Leu Cys Phe Val Pro Phe His Ile Thr Arg Ser Phe Tyr
                275                 280                 285
Leu Thr Ile Cys Phe Leu Leu Ser Gln Asp Cys Gln Leu Leu Met
                290                 295                 300
Ala Pro Ser Val Ala Tyr Lys Ile Trp Arg Pro Leu Val Ser Val
                305                 310                 315
Ser Ser Cys Leu Asn Pro Val Leu Tyr Phe Leu Ser Arg Gly Ala
                320                 325                 330
Lys Ile Glu Ser Gly Ser Ser Arg Asn
                335
```

<210> 2
<211> 335
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474840CD1

<400> 2

```
Met Thr Pro Gly Gly Arg Ala Cys Ser Glu Met Arg Ser Cys His
  1               5                  10                  15
Cys Ala Pro Ala Trp Ala Thr Glu Arg Asp Ser Val Ser Lys Lys
                20                  25                  30
Lys Lys Asn Lys Lys Lys Asn Leu Phe Ser Gln Ala Thr Ile Gly
                35                  40                  45
Leu Leu Ala Asn Thr Phe Phe Leu Phe Phe Asn Ile Phe Ile Phe
                50                  55                  60
Leu Gln Asp Gln Lys Ser Lys Pro His Asp Leu Ile Ser Cys Asn
                65                  70                  75
Ser Ala Phe Ile His Val Val Met Phe Leu Thr Val Val Asp Ala
                80                  85                  90
Trp Pro Pro Asp Met Pro Glu Ser Leu His Leu Gly Asn Glu Phe
                95                 100                 105
Lys Phe Lys Ser Leu Ser Tyr Ile Asn Arg Val Arg Met Gly Leu
                110                 115                 120
Cys Ile Cys Asn Ile Cys Leu Leu Ser Ile His Gln Ala Asn Thr
                125                 130                 135
Ile Ser Pro Asn Asn Phe Cys Leu Ala Arg Leu Lys Gln Lys Phe
                140                 145                 150
Thr Asn Asn Ile Ile Met Ser Ser Phe Phe Ser Phe Phe Phe Trp
```

```
                155                 160                 165
Ser Ile Asn Leu Ser Phe Ser Tyr Asn Ile Val Phe Phe Thr Val
                170                 175                 180
Ala Ser Ser Asn Val Thr Gln Asn Ser Leu Pro Lys Gly Ser Asn
                185                 190                 195
Thr Val His Phe Leu Pro Met Lys Ser Phe Met Arg Lys Val Phe
                200                 205                 210
Phe Thr Leu Thr Leu Ser Arg Asp Val Phe Ile Ile Gly Ile Thr
                215                 220                 225
Leu His Ser Ile Ala His Met Val Ile Leu Val Ser Arg His Glu
                230                 235                 240
Thr Gln Ser Gln His Leu His Ser Ile Ser Ile Ser Pro Gln Ala
                245                 250                 255
Phe Pro Glu Lys Arg Ala Ala Gln Thr Ile Pro Leu Leu Val Ser
                260                 265                 270
Tyr Cys Leu Val Met Cys Trp Val Asp Leu Ile Ile Ser Ser Ser
                275                 280                 285
Ser Thr Leu Leu Trp Thr Cys Asn Pro Val Phe Leu Ser Met Gln
                290                 295                 300
Asn Leu Val Gly Asp Val Tyr Ala Thr Val Val Leu Leu Glu Gln
                305                 310                 315
Ile Ser Ser Asp Lys Asn Ile Val Asp Ile Leu Gln Asn Met Gln
                320                 325                 330
Ser Ala Ile Lys Leu
                335

<210> 3
<211> 428
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475092CD1

<400> 3
Met Gln Arg Lys Glu Lys Ala Lys Cys Pro Gln Glu Ala Pro Ala
  1                 5                  10                  15
Gly Arg Glu Pro Ser Thr Pro Gly Gly Gly Ser Gly Gly Gly Gly
                 20                  25                  30
Ala Val Ala Ala Ala Ser Gly Ala Ala Val Pro Gly Ser Val Gln
                 35                  40                  45
Leu Ala Leu Ser Val Leu His Ala Leu Leu Tyr Ala Ala Leu Phe
                 50                  55                  60
Ala Phe Ala Tyr Leu Gln Leu Trp Arg Leu Leu Leu Tyr Arg Glu
                 65                  70                  75
Arg Arg Leu Ser Tyr Gln Ser Leu Cys Leu Phe Leu Cys Leu Leu
                 80                  85                  90
Trp Ala Ala Leu Arg Thr Thr Leu Phe Ser Ala Ala Phe Ser Leu
                 95                 100                 105
Ser Gly Ser Leu Pro Leu Leu Arg Pro Pro Ala His Leu His Phe
                110                 115                 120
Phe Pro His Trp Leu Leu Tyr Cys Phe Pro Ser Cys Leu Gln Phe
                125                 130                 135
Ser Thr Leu Cys Leu Leu Asn Leu Tyr Leu Ala Glu Val Ile Cys
                140                 145                 150
Lys Val Arg Cys Ala Thr Glu Leu Asp Arg His Lys Ile Leu Leu
                155                 160                 165
His Leu Gly Phe Ile Met Ala Ser Leu Leu Phe Leu Val Val Asn
                170                 175                 180
Leu Thr Cys Ala Met Leu Val His Gly Asp Val Pro Glu Asn Gln
                185                 190                 195
Leu Lys Trp Thr Val Phe Val Arg Ala Leu Ile Asn Asp Ser Leu
```

```
                200                 205                 210
Phe Ile Leu Cys Ala Ile Ser Leu Val Cys Tyr Ile Cys Lys Ile
                215                 220                 225
Thr Lys Met Ser Ser Ala Asn Val Tyr Leu Glu Ser Lys Gly Met
                230                 235                 240
Ser Leu Cys Gln Thr Val Val Val Gly Ser Val Val Ile Leu Leu
                245                 250                 255
Tyr Ser Ser Arg Ala Cys Tyr Asn Leu Val Val Val Thr Ile Ser
                260                 265                 270
Gln Asp Thr Leu Glu Ser Pro Phe Asn Tyr Gly Trp Asp Asn Leu
                275                 280                 285
Ser Asp Lys Ala His Val Glu Asp Ile Ser Gly Glu Glu Tyr Ile
                290                 295                 300
Val Phe Gly Met Val Leu Phe Leu Trp Glu His Val Pro Ala Trp
                305                 310                 315
Ser Val Val Leu Phe Phe Arg Ala Gln Arg Leu Asn Gln Asn Leu
                320                 325                 330
Ala Pro Ala Gly Met Ile Asn Ser His Ser Tyr Ser Ser Arg Ala
                335                 340                 345
Tyr Phe Phe Asp Asn Pro Arg Arg Tyr Asp Ser Asp Asp Asp Leu
                350                 355                 360
Pro Arg Leu Gly Ser Ser Arg Glu Gly Ser Leu Pro Asn Ser Gln
                365                 370                 375
Ser Leu Gly Trp Tyr Gly Thr Met Thr Gly Cys Gly Ser Ser Ser
                380                 385                 390
Tyr Thr Val Thr Pro His Leu Asn Gly Pro Met Thr Asp Thr Ala
                395                 400                 405
Pro Leu Leu Phe Thr Cys Ser Asn Leu Asp Leu Asn Asn His His
                410                 415                 420
Ser Leu Tyr Val Thr Pro Gln Asn
                425
```

<210> 4
<211> 330
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7341260CD1

<400> 4

```
Met Thr Pro Asn Ser Thr Gly Glu Val Pro Ser Pro Ile Pro Lys
  1                 5                  10                  15
Gly Ala Leu Gly Leu Ser Leu Ala Leu Ala Ser Leu Ile Ile Thr
                 20                  25                  30
Ala Asn Leu Leu Leu Ala Leu Gly Ile Ala Trp Asp Arg Arg Leu
                 35                  40                  45
Arg Ser Pro Pro Ala Gly Cys Phe Phe Leu Ser Leu Leu Leu Ala
                 50                  55                  60
Gly Leu Leu Thr Gly Leu Ala Leu Pro Thr Leu Pro Gly Leu Trp
                 65                  70                  75
Asn Gln Ser Arg Arg Gly Tyr Trp Ser Cys Leu Leu Val Tyr Leu
                 80                  85                  90
Ala Pro Asn Phe Ser Phe Leu Ser Leu Leu Ala Asn Leu Leu Leu
                 95                 100                 105
Val His Gly Glu Arg Tyr Met Ala Val Leu Arg Pro Leu Gln Pro
                110                 115                 120
Pro Gly Ser Ile Arg Leu Ala Leu Leu Leu Thr Trp Ala Gly Pro
                125                 130                 135
Leu Leu Phe Ala Ser Leu Pro Ala Leu Gly Trp Asn His Trp Thr
                140                 145                 150
Pro Gly Ala Asn Cys Ser Ser Gln Ala Ile Phe Pro Ala Pro Tyr
```

```
                155                 160                 165
Leu Tyr Leu Glu Val Tyr Gly Leu Leu Leu Pro Ala Val Gly Ala
                170                 175                 180
Ala Ala Phe Leu Ser Val Arg Val Leu Ala Thr Ala His Arg Gln
                185                 190                 195
Leu Gln Asp Ile Cys Arg Leu Glu Arg Ala Val Cys Arg Asp Glu
                200                 205                 210
Pro Ser Ala Leu Ala Arg Ala Leu Thr Trp Arg Gln Ala Arg Ala
                215                 220                 225
Gln Ala Gly Ala Met Leu Leu Phe Gly Leu Cys Trp Gly Pro Tyr
                230                 235                 240
Val Ala Thr Leu Leu Leu Ser Val Leu Ala Tyr Glu Gln Arg Pro
                245                 250                 255
Pro Leu Gly Pro Gly Thr Leu Leu Ser Leu Leu Ser Leu Gly Ser
                260                 265                 270
Ala Ser Ala Ala Ala Val Pro Val Ala Met Gly Leu Gly Asp Gln
                275                 280                 285
Arg Tyr Thr Ala Pro Trp Arg Ala Ala Ala Gln Arg Cys Leu Gln
                290                 295                 300
Gly Leu Trp Gly Arg Ala Ser Arg Asp Ser Pro Gly Pro Ser Ile
                305                 310                 315
Ala Tyr His Pro Ser Ser Gln Ser Ser Val Asp Leu Asp Leu Asn
                320                 325                 330
```

<210> 5
<211> 676
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473911CD1

<400> 5

```
Met Asn Lys Asn Asn Lys Pro Ser Ser Phe Ile Ala Ile Arg Asn
  1               5                  10                  15
Ala Ala Phe Ser Glu Val Gly Ile Gly Ile Ser Ala Asn Ala Met
                20                  25                  30
Leu Leu Leu Phe His Ile Leu Thr Cys Leu Leu Lys His Arg Thr
                35                  40                  45
Lys Pro Ala Asp Leu Ile Val Cys His Val Ala Leu Ile His Ile
                50                  55                  60
Ile Leu Leu Leu Pro Thr Glu Phe Ile Ala Thr Asp Ile Phe Gly
                65                  70                  75
Ser Gln Asp Ser Glu Asp Asp Ile Lys His Lys Ser Val Ile Tyr
                80                  85                  90
Arg Arg Asn Arg Gln Ser Gln His Phe His Ser Thr Asn Leu Ser
                95                 100                 105
Pro Lys Ala Pro Pro Glu Lys Met Ala Thr Gln Thr Ile Leu Leu
                110                 115                 120
Leu Val Ser Cys Phe Val Ile Val Tyr Val Leu Asp Cys Val Val
                125                 130                 135
Ala Ser Cys Ser Gly Leu Val Trp Asn Ser Asp Pro Val Arg His
                140                 145                 150
Arg Val Gln Met Leu Val Asp Asn Gly Tyr Ala Thr Ile Ser Pro
                155                 160                 165
Ser Val Leu Pro Arg Leu Thr Ala Pro Asn Glu Trp Arg Ala Ser
                170                 175                 180
Val Tyr Leu Asn Asp Ser Leu Asn Lys Cys Ser Asn Gly Arg Leu
                185                 190                 195
Leu Cys Val Asp Arg Gly Leu Asp Glu Gly Pro Arg Ser Val Pro
                200                 205                 210
```

```
Lys Cys Ser Glu Ser Glu Thr Asp Glu Asp Tyr Ile Val Leu Arg
                215                 220                 225
Ala Pro Leu Arg Glu Asp Glu Pro Lys Asp Gly Gly Ser Val Gly
                230                 235                 240
Asn Ala Ala Leu Val Ser Pro Glu Ala Ser Ala Glu Glu Glu Glu
                245                 250                 255
Glu Arg Glu Glu Gly Gly Glu Ala Cys Gly Leu Glu Arg Thr Gly
                260                 265                 270
Ala Gly Gly Glu Gln Val Asp Leu Gly Glu Leu Pro Asp His Glu
                275                 280                 285
Glu Lys Ser Asn Gln Lys Val Ala Ala Ala Thr Leu Glu Asp Arg
                290                 295                 300
Thr Gln Asp Glu Pro Ala Glu Glu Ser Cys Gln Ile Val Leu Phe
                305                 310                 315
Gln Asn Asn Cys Met Asp Asn Phe Val Thr Ser Leu Thr Gly Ser
                320                 325                 330
Pro Tyr Glu Phe Phe Pro Thr Lys Ser Thr Ser Phe Cys Arg Glu
                335                 340                 345
Ser Cys Ser Pro Phe Ser Glu Ser Val Lys Ser Leu Glu Ser Glu
                350                 355                 360
Gln Ala Pro Lys Leu Gly Leu Cys Ala Glu Glu Asp Pro Val Val
                365                 370                 375
Gly Ala Leu Cys Gly Gln His Gly Pro Leu Gln Asp Gly Val Ala
                380                 385                 390
Glu Gly Pro Thr Ala Pro Asp Val Val Val Leu Pro Lys Glu Glu
                395                 400                 405
Glu Lys Glu Glu Val Ile Val Asp Asp Met Leu Ala Asn Pro Tyr
                410                 415                 420
Val Met Gly Asp Glu Gly Glu Glu Glu Glu Glu Glu Phe Val Asp
                425                 430                 435
Asp Thr Leu Ala Asn Pro Tyr Val Met Gly Val Gly Leu Pro Gly
                440                 445                 450
Arg Gly Gly Glu Glu Glu Glu Glu Glu Glu Val Val Asp Asp Thr
                455                 460                 465
Leu Ala Ser Leu Tyr Lys Met Gly Glu Glu His Arg His Lys Gly
                470                 475                 480
Leu Ala Pro Leu Trp Glu Gly Gly Gln Lys Pro Ser Gln Lys Leu
                485                 490                 495
Pro Pro Lys Lys Pro Asp Leu Arg Gln Val Pro Gln Pro Leu Ala
                500                 505                 510
Ser Glu Val Pro Gln Arg Arg Gln Glu Arg Ala Val Val Thr Glu
                515                 520                 525
Gly Arg Pro Leu Glu Ala Ser Arg Ala Leu Pro Ala Lys Pro Arg
                530                 535                 540
Ala Phe Thr Leu Tyr Pro Arg Ser Phe Ser Val Glu Gly Gln Glu
                545                 550                 555
Ile Pro Val Ser Ile Ser Val Tyr Trp Glu Pro Glu Gly Ser Gly
                560                 565                 570
Leu Asp Asp His Arg Ile Lys Arg Lys Glu Glu His Leu Ser Val
                575                 580                 585
Val Ser Gly Ser Phe Ser Gln Arg Asn His Leu Pro Ser Ser Gly
                590                 595                 600
Thr Ser Thr Pro Ser Ser Met Val Asp Ile Pro Pro Pro Phe Asp
                605                 610                 615
Leu Ala Cys Ile Thr Lys Lys Pro Ile Thr Lys Ser Ser Pro Ser
                620                 625                 630
Leu Leu Ile Asp Ser Asp Ser Pro Asp Lys Tyr Lys Lys Lys Lys
                635                 640                 645
Ser Ser Phe Lys Arg Phe Leu Ala Leu Met Phe Asn Lys Met Glu
                650                 655                 660
Arg Pro Gly Thr Met Ala His Ala Cys His Pro Ser Thr Leu Gly
                665                 670                 675
Ser
```

<210> 6
<211> 372
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474767CD1

<400> 6

```
Met Glu His Thr His Ala His Leu Ala Ala Asn Ser Ser Leu Ser
  1               5                  10                  15
Trp Trp Ser Pro Gly Ser Ala Cys Gly Leu Gly Phe Val Pro Val
             20                  25                  30
Val Tyr Tyr Ser Leu Leu Leu Cys Leu Gly Leu Pro Ala Asn Ile
             35                  40                  45
Leu Thr Val Ile Ile Leu Ser Gln Leu Val Ala Arg Arg Gln Lys
             50                  55                  60
Ser Ser Tyr Asn Tyr Leu Leu Ala Leu Ala Ala Ala Asp Ile Leu
             65                  70                  75
Val Leu Phe Phe Ile Val Phe Val Asp Phe Leu Leu Glu Asp Phe
             80                  85                  90
Ile Leu Asn Met Gln Met Pro Gln Val Pro Asp Lys Ile Ile Glu
             95                 100                 105
Val Leu Glu Phe Ser Ser Ile His Thr Ser Ile Trp Ile Thr Val
            110                 115                 120
Pro Leu Thr Ile Asp Arg Tyr Ile Ala Val Cys His Pro Leu Lys
            125                 130                 135
Tyr His Thr Val Ser Tyr Pro Ala Arg Thr Arg Lys Val Ile Val
            140                 145                 150
Ser Val Tyr Ile Thr Cys Phe Leu Thr Ser Ile Pro Tyr Tyr Trp
            155                 160                 165
Trp Pro Asn Ile Trp Thr Glu Asp Tyr Ile Ser Thr Ser Val His
            170                 175                 180
His Val Leu Ile Trp Ile His Cys Phe Thr Val Tyr Leu Val Pro
            185                 190                 195
Cys Ser Ile Phe Phe Ile Leu Asn Ser Ile Ile Val Tyr Lys Leu
            200                 205                 210
Arg Arg Lys Ser Asn Phe Arg Leu Arg Gly Tyr Ser Thr Gly Lys
            215                 220                 225
Thr Thr Ala Ile Leu Phe Thr Ile Thr Ser Ile Phe Ala Thr Leu
            230                 235                 240
Trp Ala Pro Arg Ile Ile Met Ile Leu Tyr His Leu Tyr Gly Ala
            245                 250                 255
Pro Ile Gln Asn Arg Trp Leu Val His Ile Met Ser Asp Ile Ala
            260                 265                 270
Asn Met Leu Ala Leu Leu Asn Thr Ala Ile Asn Phe Phe Leu Tyr
            275                 280                 285
Cys Phe Ile Ser Lys Arg Phe Arg Thr Met Ala Ala Ala Thr Leu
            290                 295                 300
Lys Ala Phe Phe Lys Cys Gln Lys Gln Pro Val Gln Phe Tyr Thr
            305                 310                 315
Asn His Asn Phe Ser Ile Thr Ser Ser Pro Trp Ile Ser Pro Ala
            320                 325                 330
Asn Ser His Cys Ile Lys Met Leu Val Tyr Gln Tyr Asp Lys Asn
            335                 340                 345
Gly Lys Pro Ile Lys Ser Arg Asn Asp Ser Lys Ser Ser Tyr Gln
            350                 355                 360
Phe Glu Asp Ala Ile Gly Ala Cys Val Ile Ile Leu
            365                 370
```

<210> 7
<211> 271
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475815CD1

<400> 7
Met Asn Lys Asn Asn Lys Pro Ser Ser Phe Ile Ala Ile Arg Asn
1 5 10 15
Ala Ala Phe Ser Glu Val Gly Ile Gly Ile Ser Ala Asn Ala Met
20 25 30
Leu Leu Leu Phe His Ile Leu Thr Cys Leu Leu Lys His Arg Thr
35 40 45
Lys Pro Ala Asp Leu Ile Val Cys His Val Ala Leu Ile His Ile
50 55 60
Ile Leu Leu Leu Pro Thr Glu Phe Ile Ala Thr Asp Ile Phe Gly
65 70 75
Ser Gln Asp Ser Glu Asp Asp Ile Lys His Lys Ser Val Ile Tyr
80 85 90
Arg Tyr Arg Leu Met Arg Gly Leu Ser Ile Ser Thr Thr Cys Leu
95 100 105
Leu Ser Ile Leu Pro Ala Ile Thr Cys Ser Pro Arg Ser Ser Cys
110 115 120
Leu Ala Val Phe Lys Asp Ser His Ile Thr Asn His Val Ala Phe
125 130 135
Ser Ser Val Phe His Ile Ser Ile Ser Asp Ser Phe Leu Val Ser
140 145 150
Thr Leu Pro Ile Lys Asn Leu Ala Ser Asn Ser Leu Thr Phe Val
155 160 165
Thr Gln Ser Cys Ser Ala Gly Ile Gly Ser Arg Pro Pro Ser Ser
170 175 180
Gly Tyr Met Val Ile Leu Leu Ser Arg Arg Asn Arg Gln Ser Gln
185 190 195
His Phe His Ser Thr Asn Leu Ser Pro Lys Ala Pro Pro Glu Lys
200 205 210
Met Ala Thr Gln Thr Ile Leu Leu Leu Val Ser Cys Phe Val Ile
215 220 225
Val Tyr Val Leu Asp Cys Val Val Ala Ser Cys Ser Gly Leu Val
230 235 240
Trp Asn Ser Asp Pro Val Arg His Arg Val Gln Met Leu Val Asp
245 250 255
Asn Gly Tyr Ala Thr Ile Ser Pro Ser Val Leu Val Ser Thr Glu
260 265 270
Lys

<210> 8
<211> 611
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 60263275CD1

<400> 8
Met Gln Gly Pro Leu Leu Leu Pro Gly Leu Cys Phe Leu Leu Ser
1 5 10 15
Leu Phe Gly Ala Val Thr Gln Lys Thr Lys Asn Ile Asn Glu Cys
20 25 30

```
Thr Pro Pro Tyr Ser Val Tyr Cys Gly Phe Asn Ala Val Cys Tyr
                35                  40                  45
Asn Val Glu Gly Ser Phe Tyr Cys Gln Cys Val Pro Gly Tyr Arg
                50                  55                  60
Leu His Ser Gly Asn Glu Gln Phe Ser Asn Ser Asn Glu Asn Thr
                65                  70                  75
Cys Gln Asp Thr Thr Ser Ser Lys Thr Thr Gln Gly Arg Lys Glu
                80                  85                  90
Leu Gln Lys Ile Val Asp Lys Phe Glu Ser Leu Leu Thr Asn Gln
                95                 100                 105
Thr Leu Trp Arg Thr Glu Gly Arg Gln Glu Ile Ser Ser Thr Ala
               110                 115                 120
Thr Thr Ile Leu Arg Asp Val Glu Ser Lys Val Leu Glu Thr Ala
               125                 130                 135
Leu Lys Asp Pro Glu Gln Lys Val Leu Lys Ile Gln Asn Asp Ser
               140                 145                 150
Val Ala Ile Glu Thr Gln Ala Ile Thr Asp Asn Cys Ser Glu Glu
               155                 160                 165
Arg Lys Thr Phe Asn Leu Asn Val Gln Met Asn Ser Met Asp Ile
               170                 175                 180
Arg Cys Ser Asp Ile Ile Gln Gly Asp Thr Gln Gly Pro Ser Ala
               185                 190                 195
Ile Ala Phe Ile Ser Tyr Ser Ser Leu Gly Asn Ile Ile Asn Ala
               200                 205                 210
Thr Phe Phe Glu Glu Met Asp Lys Lys Asp Gln Val Tyr Leu Asn
               215                 220                 225
Ser Gln Val Val Ser Ala Ala Ile Gly Pro Lys Arg Asn Val Ser
               230                 235                 240
Leu Ser Lys Ser Val Thr Leu Thr Phe Gln His Val Lys Met Thr
               245                 250                 255
Pro Ser Thr Lys Lys Val Phe Cys Val Tyr Trp Lys Ser Thr Gly
               260                 265                 270
Gln Gly Ser Gln Trp Ser Arg Asp Gly Cys Phe Leu Ile His Val
               275                 280                 285
Asn Lys Ser His Thr Met Cys Asn Cys Ser His Leu Ser Ser Phe
               290                 295                 300
Ala Val Leu Met Ala Leu Thr Ser Gln Glu Glu Asp Pro Val Leu
               305                 310                 315
Thr Val Ile Thr Tyr Val Gly Leu Ser Val Ser Leu Leu Cys Leu
               320                 325                 330
Leu Leu Ala Ala Leu Thr Phe Leu Leu Cys Lys Ala Ile Gln Asn
               335                 340                 345
Thr Ser Thr Ser Leu His Leu Gln Leu Ser Leu Cys Leu Phe Leu
               350                 355                 360
Ala His Leu Leu Phe Leu Val Gly Ile Asp Arg Thr Glu Pro Lys
               365                 370                 375
Val Leu Cys Ser Ile Ile Ala Gly Ala Leu His Tyr Leu Tyr Leu
               380                 385                 390
Ala Ala Phe Thr Trp Met Leu Leu Glu Gly Val His Leu Phe Leu
               395                 400                 405
Thr Ala Arg Asn Leu Thr Val Val Asn Tyr Ser Ser Ile Asn Arg
               410                 415                 420
Leu Met Lys Trp Ile Met Phe Pro Val Gly Tyr Gly Val Pro Ala
               425                 430                 435
Val Thr Val Ala Ile Ser Ala Ala Ser Trp Pro His Leu Tyr Gly
               440                 445                 450
Thr Ala Asp Arg Cys Trp Leu His Leu Asp Gln Gly Phe Met Trp
               455                 460                 465
Ser Phe Leu Gly Pro Val Cys Ala Ile Phe Ser Ala Asn Leu Val
               470                 475                 480
Leu Phe Ile Leu Val Phe Trp Ile Leu Lys Arg Lys Leu Ser Ser
               485                 490                 495
Leu Asn Ser Glu Val Ser Thr Ile Gln Asn Thr Arg Met Leu Ala
```

```
                500                 505                 510
Phe Lys Ala Thr Ala Gln Leu Phe Ile Leu Gly Cys Thr Trp Cys
                515                 520                 525
Leu Gly Leu Leu Gln Val Gly Pro Ala Ala Gln Val Met Ala Tyr
                530                 535                 540
Leu Phe Thr Ile Ile Asn Ser Leu Gln Gly Phe Phe Ile Phe Leu
                545                 550                 555
Val Tyr Cys Leu Leu Ser Gln Gln Val Gln Lys Gln Tyr Gln Lys
                560                 565                 570
Trp Phe Arg Glu Ile Val Lys Ser Lys Ser Glu Ser Glu Thr Tyr
                575                 580                 585
Thr Leu Ser Ser Lys Met Gly Pro Asp Ser Lys Pro Ser Glu Gly
                590                 595                 600
Asp Val Phe Pro Gly Gln Val Lys Arg Lys Tyr
                605                 610
```

<210> 9
<211> 1469
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 60203310CD1

<400> 9

```
Met Trp Pro Ser Gln Leu Leu Ile Phe Met Met Leu Leu Ala Pro
  1                 5                  10                  15
Ile Ile His Ala Phe Ser Arg Ala Pro Ile Pro Met Ala Val Val
                 20                  25                  30
Arg Arg Glu Leu Ser Cys Glu Ser Tyr Pro Ile Glu Leu Arg Cys
                 35                  40                  45
Pro Gly Thr Asp Val Ile Met Ile Glu Ser Ala Asn Tyr Gly Arg
                 50                  55                  60
Thr Asp Asp Lys Ile Cys Asp Ser Asp Pro Ala Gln Met Glu Asn
                 65                  70                  75
Ile Arg Cys Tyr Leu Pro Asp Ala Tyr Lys Ile Met Ser Gln Arg
                 80                  85                  90
Cys Asn Asn Arg Thr Gln Cys Ala Val Val Ala Gly Pro Asp Val
                 95                 100                 105
Phe Pro Asp Pro Cys Pro Gly Thr Tyr Lys Tyr Leu Glu Val Gln
                110                 115                 120
Tyr Glu Cys Val Pro Tyr Lys Val Glu Gln Lys Val Phe Leu Cys
                125                 130                 135
Pro Gly Leu Leu Lys Gly Val Tyr Gln Ser Glu His Leu Phe Glu
                140                 145                 150
Ser Asp His Gln Ser Gly Ala Trp Cys Lys Asp Pro Leu Gln Ala
                155                 160                 165
Ser Asp Lys Ile Tyr Tyr Met Pro Trp Thr Pro Tyr Arg Thr Asp
                170                 175                 180
Thr Leu Thr Glu Tyr Ser Ser Lys Asp Asp Phe Ile Ala Gly Arg
                185                 190                 195
Pro Thr Thr Thr Tyr Lys Leu Pro His Arg Val Asp Gly Thr Gly
                200                 205                 210
Phe Val Val Tyr Asp Gly Ala Leu Phe Phe Asn Lys Glu Arg Thr
                215                 220                 225
Arg Asn Ile Val Lys Phe Asp Leu Arg Thr Arg Ile Lys Ser Gly
                230                 235                 240
Glu Ala Ile Ile Ala Asn Ala Asn Tyr His Asp Thr Ser Pro Tyr
                245                 250                 255
Arg Trp Gly Gly Lys Ser Asp Ile Asp Leu Ala Val Asp Glu Asn
                260                 265                 270
Gly Leu Trp Val Ile Tyr Ala Thr Glu Gln Asn Asn Gly Lys Ile
```

```
                275                 280                 285
Val Ile Ser Gln Leu Asn Pro Tyr Thr Leu Arg Ile Glu Gly Thr
            290                 295                 300
Trp Asp Thr Ala Tyr Asp Lys Arg Ser Ala Ser Asn Ala Phe Met
        305                 310                 315
Ile Cys Gly Ile Leu Tyr Val Val Lys Ser Val Tyr Glu Asp Asp
                320                 325                 330
Asp Asn Glu Ala Thr Gly Asn Lys Ile Asp Tyr Ile Tyr Asn Thr
            335                 340                 345
Asp Gln Ser Lys Asp Ser Leu Val Asp Val Pro Phe Pro Asn Ser
        350                 355                 360
Tyr Gln Tyr Ile Ala Ala Val Asp Tyr Asn Pro Arg Asp Asn Leu
    365                 370                 375
Leu Tyr Val Trp Asn Asn Tyr His Val Val Lys Tyr Ser Leu Asp
380                 385                 390
Phe Gly Pro Leu Asp Ser Arg Ser Gly Gln Ala His His Gly Gln
            395                 400                 405
Val Ser Tyr Ile Ser Pro Pro Ile His Leu Asp Ser Glu Leu Glu
        410                 415                 420
Arg Pro Ser Val Lys Asp Ile Ser Thr Thr Gly Pro Leu Gly Met
    425                 430                 435
Gly Ser Thr Thr Thr Ser Thr Thr Leu Arg Thr Thr Thr Leu Ser
440                 445                 450
Pro Gly Arg Ser Thr Thr Pro Ser Val Ser Gly Arg Arg Asn Arg
            455                 460                 465
Ser Thr Ser Thr Pro Ser Pro Ala Val Glu Val Leu Asp Asp Met
        470                 475                 480
Thr Thr His Leu Pro Ser Ala Ser Ser Gln Ile Pro Ala Leu Glu
    485                 490                 495
Glu Ser Cys Glu Ala Val Glu Ala Arg Glu Ile Met Trp Phe Lys
500                 505                 510
Thr Arg Gln Gly Gln Ile Ala Lys Gln Pro Cys Pro Ala Gly Thr
            515                 520                 525
Ile Gly Val Ser Thr Tyr Leu Cys Leu Ala Pro Asp Gly Ile Trp
        530                 535                 540
Asp Pro Gln Gly Pro Asp Leu Ser Asn Cys Ser Ser Pro Trp Val
    545                 550                 555
Asn His Ile Thr Gln Lys Leu Lys Ser Gly Glu Thr Ala Ala Asn
560                 565                 570
Ile Ala Arg Glu Leu Ala Glu Gln Thr Arg Asn His Leu Asn Ala
            575                 580                 585
Gly Asp Ile Thr Tyr Ser Val Arg Ala Met Asp Gln Leu Val Gly
        590                 595                 600
Leu Leu Asp Val Gln Leu Arg Asn Leu Thr Pro Gly Gly Lys Asp
    605                 610                 615
Ser Ala Ala Arg Ser Leu Asn Lys Leu Gln Lys Arg Glu Arg Ser
620                 625                 630
Cys Arg Ala Tyr Val Gln Ala Met Val Glu Thr Val Asn Asn Leu
            635                 640                 645
Leu Gln Pro Gln Ala Leu Asn Ala Trp Arg Asp Leu Thr Thr Ser
        650                 655                 660
Asp Gln Leu Arg Ala Ala Thr Met Leu Leu His Thr Val Glu Glu
    665                 670                 675
Ser Ala Phe Val Leu Ala Asp Asn Leu Leu Lys Thr Asp Ile Val
680                 685                 690
Arg Glu Asn Thr Asp Asn Ile Lys Leu Glu Val Ala Arg Leu Ser
            695                 700                 705
Thr Glu Gly Asn Leu Glu Asp Leu Lys Phe Pro Glu Asn Met Gly
        710                 715                 720
His Gly Ser Thr Ile Gln Leu Ser Ala Asn Thr Leu Lys Gln Asn
    725                 730                 735
Gly Arg Asn Gly Glu Ile Arg Val Ala Phe Val Leu Tyr Asn Asn
740                 745                 750
```

```
Leu Gly Pro Tyr Leu Ser Thr Glu Asn Ala Ser Met Lys Leu Gly
                755                 760                 765
Thr Glu Ala Leu Ser Thr Asn His Ser Val Ile Val Asn Ser Pro
                770                 775                 780
Val Ile Thr Ala Ala Ile Asn Lys Glu Phe Ser Asn Lys Val Tyr
                785                 790                 795
Leu Ala Asp Pro Val Val Phe Thr Val Lys His Ile Lys Gln Ser
                800                 805                 810
Glu Glu Asn Phe Asn Pro Asn Cys Ser Phe Trp Ser Tyr Ser Lys
                815                 820                 825
Arg Thr Met Thr Gly Tyr Trp Ser Thr Gln Gly Cys Arg Leu Leu
                830                 835                 840
Thr Thr Asn Lys Thr His Thr Thr Cys Ser Cys Asn His Leu Thr
                845                 850                 855
Asn Phe Ala Val Leu Met Ala His Val Glu Val Lys His Ser Asp
                860                 865                 870
Ala Val His Asp Leu Leu Leu Asp Val Ile Thr Trp Val Gly Ile
                875                 880                 885
Leu Leu Ser Leu Val Cys Leu Leu Ile Cys Ile Phe Thr Phe Cys
                890                 895                 900
Phe Phe Arg Gly Leu Gln Ser Asp Arg Asn Thr Ile His Lys Asn
                905                 910                 915
Leu Cys Ile Ser Leu Phe Val Ala Glu Leu Leu Phe Leu Ile Gly
                920                 925                 930
Ile Asn Arg Thr Asp Gln Pro Ile Ala Cys Ala Val Phe Ala Ala
                935                 940                 945
Leu Leu His Phe Phe Phe Leu Ala Ala Phe Thr Trp Met Phe Leu
                950                 955                 960
Glu Gly Val Gln Leu Tyr Ile Met Leu Val Glu Val Phe Glu Ser
                965                 970                 975
Glu His Ser Arg Arg Lys Tyr Phe Tyr Leu Val Gly Tyr Gly Met
                980                 985                 990
Pro Ala Leu Ile Val Ala Val Ser Ala Ala Val Asp Tyr Arg Ser
                995                1000                1005
Tyr Gly Thr Asp Lys Val Cys Trp Leu Arg Leu Asp Thr Tyr Phe
               1010                1015                1020
Ile Trp Ser Phe Ile Gly Pro Ala Thr Leu Ile Ile Met Leu Asn
               1025                1030                1035
Val Ile Phe Leu Gly Ile Ala Leu Tyr Lys Met Val His His Thr
               1040                1045                1050
Ala Ile Leu Lys Pro Glu Ser Gly Cys Leu Asp Asn Ile Asn Tyr
               1055                1060                1065
Glu Asp Asn Arg Pro Phe Ile Lys Ser Trp Val Ile Gly Ala Ile
               1070                1075                1080
Ala Leu Leu Cys Leu Leu Gly Leu Thr Trp Ala Phe Gly Leu Met
               1085                1090                1095
Tyr Ile Asn Glu Ser Thr Val Ile Met Ala Tyr Leu Phe Thr Ile
               1100                1105                1110
Phe Asn Ser Leu Gln Gly Met Phe Ile Phe Ile Phe His Cys Val
               1115                1120                1125
Leu Gln Lys Lys Val Arg Lys Glu Tyr Gly Lys Cys Leu Arg Thr
               1130                1135                1140
His Cys Cys Ser Gly Lys Ser Thr Glu Ser Ser Ile Gly Ser Gly
               1145                1150                1155
Lys Thr Ser Gly Ser Arg Thr Pro Gly Arg Tyr Ser Thr Gly Ser
               1160                1165                1170
Gln Ser Arg Ile Arg Arg Met Trp Asn Asp Thr Val Arg Lys Gln
               1175                1180                1185
Ser Glu Ser Ser Phe Ile Thr Gly Asp Ile Asn Ser Ser Ala Ser
               1190                1195                1200
Leu Asn Arg Glu Gly Leu Leu Asn Asn Ala Arg Asp Thr Ser Val
               1205                1210                1215
Met Asp Thr Leu Pro Leu Asn Gly Asn His Gly Asn Ser Tyr Ser
```

```
                1220                1225                1230
Ile Ala Ser Gly Glu Tyr Leu Ser Asn Cys Val Gln Ile Ile Asp
                1235                1240                1245
Arg Gly Tyr Asn His Asn Glu Thr Ala Leu Glu Lys Lys Ile Leu
                1250                1255                1260
Lys Glu Leu Thr Ser Asn Tyr Ile Pro Ser Tyr Leu Asn Asn His
                1265                1270                1275
Glu Arg Ser Ser Glu Gln Asn Arg Asn Leu Met Asn Lys Leu Val
                1280                1285                1290
Asn Asn Leu Gly Ser Gly Arg Glu Asp Asp Ala Ile Val Leu Asp
                1295                1300                1305
Asp Ala Thr Ser Phe Asn His Glu Glu Ser Leu Gly Leu Glu Leu
                1310                1315                1320
Ile His Glu Glu Ser Asp Ala Pro Leu Leu Pro Pro Arg Val Tyr
                1325                1330                1335
Ser Thr Glu Asn His Gln Pro His His Tyr Thr Arg Arg Arg Ile
                1340                1345                1350
Pro Gln Asp His Ser Glu Ser Phe Phe Pro Leu Leu Thr Asn Glu
                1355                1360                1365
His Thr Glu Asp Leu Gln Ser Pro His Arg Asp Ser Leu Tyr Thr
                1370                1375                1380
Ser Met Pro Thr Leu Ala Gly Val Ala Ala Thr Glu Ser Val Thr
                1385                1390                1395
Thr Ser Thr Gln Thr Glu Pro Pro Pro Ala Lys Cys Gly Asp Ala
                1400                1405                1410
Glu Asp Val Tyr Tyr Lys Ser Met Pro Asn Leu Gly Ser Arg Asn
                1415                1420                1425
His Val His Gln Leu His Thr Tyr Tyr Gln Leu Gly Arg Gly Ser
                1430                1435                1440
Ser Asp Gly Phe Ile Val Pro Pro Asn Lys Asp Gly Thr Pro Pro
                1445                1450                1455
Glu Gly Ser Ser Lys Gly Pro Ala His Leu Val Thr Ser Leu
                1460                1465
```

<210> 10
<211> 469
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477349CD1

<400> 10

```
Met Asp Pro Ser Val Val Ser Asn Glu Tyr Tyr Asp Val Ala His
  1               5                  10                  15
Gly Ala Lys Asp Pro Val Val Pro Thr Ser Leu Gln Asp Ile Thr
                20                  25                  30
Ala Val Leu Gly Thr Glu Ala Tyr Thr Glu Glu Asp Lys Ser Met
                35                  40                  45
Val Ser His Ala Gln Lys Ser Gln His Ser Cys Leu Ser His Ser
                50                  55                  60
Arg Trp Leu Arg Ser Pro Gln Val Thr Gly Gly Ser Trp Asp Leu
                65                  70                  75
Arg Ile Arg Pro Ser Lys Asp Ser Ser Ser Phe Arg Gln Ala Gln
                80                  85                  90
Cys Leu Arg Lys Asp Pro Gly Ala Asn Asn His Leu Glu Ser Gln
                95                 100                 105
Gly Val Arg Gly Thr Ala Gly Asp Ala Asp Arg Glu Leu Arg Gly
               110                 115                 120
Pro Ser Glu Lys Ala Thr Ala Gly Gln Pro Arg Val Thr Leu Leu
               125                 130                 135
Pro Thr Pro Asn Val Ser Gly Leu Ser Gln Glu Phe Glu Ser His
```

```
                  140                 145                 150
Trp Pro Glu Ile Ala Glu Arg Ser Pro Cys Val Ala Gly Val Ile
                  155                 160                 165
Pro Val Ile Tyr Tyr Ser Val Leu Leu Gly Leu Gly Leu Pro Val
                  170                 175                 180
Ser Leu Leu Thr Ala Val Ala Leu Ala Arg Leu Ala Thr Arg Thr
                  185                 190                 195
Arg Arg Pro Ser Tyr Tyr Tyr Leu Leu Ala Leu Thr Ala Ser Asp
                  200                 205                 210
Ile Ile Ile Gln Val Val Ile Val Phe Ala Gly Phe Leu Leu Gln
                  215                 220                 225
Gly Ala Val Leu Ala Arg Gln Val Pro Gln Ala Val Val Arg Thr
                  230                 235                 240
Ala Asn Ile Leu Glu Phe Ala Ala Asn His Ala Ser Val Trp Ile
                  245                 250                 255
Ala Ile Leu Leu Thr Val Asp Arg Tyr Thr Ala Leu Cys His Pro
                  260                 265                 270
Leu His His Arg Ala Ala Ser Ser Pro Gly Arg Thr Arg Arg Ala
                  275                 280                 285
Ile Ala Ala Val Leu Ser Ala Ala Leu Leu Thr Gly Ile Pro Phe
                  290                 295                 300
Tyr Trp Trp Leu Asp Met Trp Arg Asp Thr Asp Ser Pro Arg Thr
                  305                 310                 315
Leu Asp Glu Val Leu Lys Trp Ala His Cys Leu Thr Val Tyr Phe
                  320                 325                 330
Ile Pro Cys Gly Val Phe Leu Val Thr Asn Ser Ala Ile Ile His
                  335                 340                 345
Arg Leu Arg Arg Arg Gly Arg Ser Gly Leu Gln Pro Arg Val Gly
                  350                 355                 360
Lys Ser Thr Ala Ile Leu Leu Gly Ile Thr Thr Leu Phe Thr Leu
                  365                 370                 375
Leu Trp Ala Pro Arg Val Phe Val Met Leu Tyr His Met Tyr Val
                  380                 385                 390
Ala Pro Val His Arg Asp Trp Arg Val His Leu Ala Leu Asp Val
                  395                 400                 405
Ala Asn Met Val Ala Met Leu His Thr Ala Ala Asn Phe Gly Leu
                  410                 415                 420
Tyr Cys Phe Val Ser Lys Thr Phe Arg Ala Thr Val Arg Gln Val
                  425                 430                 435
Ile His Asp Ala Tyr Leu Pro Cys Thr Leu Ala Ser Gln Pro Glu
                  440                 445                 450
Gly Met Ala Ala Lys Pro Val Met Glu Pro Pro Gly Leu Pro Thr
                  455                 460                 465
Gly Ala Glu Val
```

<210> 11
<211> 335
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55002225CD1

<400> 11

```
Met Asn Pro Phe His Ala Ser Cys Trp Asn Thr Ser Ala Glu Leu
  1               5                  10                  15
Leu Asn Lys Ser Trp Asn Lys Glu Phe Ala Tyr Gln Thr Ala Ser
                 20                  25                  30
Val Val Asp Thr Val Ile Leu Pro Ser Met Ile Gly Ile Ile Cys
                 35                  40                  45
Ser Thr Gly Leu Val Gly Asn Ile Leu Ile Val Phe Thr Ile Ile
```

```
                50                  55                  60
Arg Ser Arg Lys Lys Thr Val Pro Asp Ile Tyr Ile Cys Asn Leu
                65                  70                  75
Ala Val Ala Asp Leu Val His Ile Val Gly Met Pro Phe Leu Ile
                80                  85                  90
His Gln Trp Ala Arg Gly Gly Glu Trp Val Phe Gly Gly Pro Leu
                95                 100                 105
Cys Thr Ile Ile Thr Ser Leu Asp Thr Cys Asn Gln Phe Ala Cys
               110                 115                 120
Ser Ala Ile Met Thr Val Met Ser Val Asp Arg Tyr Phe Ala Leu
               125                 130                 135
Val Gln Pro Phe Arg Leu Thr Arg Trp Arg Thr Arg Tyr Lys Thr
               140                 145                 150
Ile Arg Ile Asn Leu Gly Leu Trp Ala Ala Ser Phe Ile Leu Ala
               155                 160                 165
Leu Pro Val Trp Val Tyr Ser Lys Val Ile Lys Phe Lys Asp Gly
               170                 175                 180
Val Glu Ser Cys Ala Phe Asp Leu Thr Ser Pro Asp Asp Val Leu
               185                 190                 195
Trp Tyr Thr Leu Tyr Leu Thr Ile Thr Thr Phe Phe Phe Pro Leu
               200                 205                 210
Pro Leu Ile Leu Val Cys Tyr Ile Leu Ile Leu Cys Tyr Thr Trp
               215                 220                 225
Glu Met Tyr Gln Gln Asn Lys Asp Ala Arg Cys Cys Asn Pro Ser
               230                 235                 240
Val Pro Lys Gln Arg Val Met Lys Leu Thr Lys Met Val Leu Val
               245                 250                 255
Leu Val Val Val Phe Ile Leu Ser Ala Ala Pro Tyr His Val Ile
               260                 265                 270
Gln Leu Val Asn Leu Gln Met Glu Gln Pro Thr Leu Ala Phe Tyr
               275                 280                 285
Val Gly Tyr Tyr Leu Ser Ile Cys Leu Ser Tyr Ala Ser Ser Ser
               290                 295                 300
Ile Asn Pro Phe Leu Tyr Ile Leu Leu Ser Gly Thr Pro Gln Ile
               305                 310                 315
Gln Arg Arg Ala Thr Glu Lys Glu Ile Asn Asn Met Gly Asn Thr
               320                 325                 330
Leu Lys Ser His Phe
               335
```

<210> 12
<211> 630
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475686CD1

<400> 12

```
Met Arg Leu Gly Pro Val Pro Ala Arg Ala Arg Ala Leu Leu Ser
  1               5                  10                  15
Trp Val Arg Gly Leu Glu Ser Arg Gly Gly Glu Trp Thr Lys Cys
                20                  25                  30
Ile Val Gln Leu Gly His Leu Leu Ala Thr Gln His Pro Ala Ala
                35                  40                  45
Pro Thr Cys Gly Val Val Ser Ser Ala Leu Val Met His Ser Thr
                50                  55                  60
Asp Val Cys Leu Ala Pro Thr Met His Gln Ala Leu Asp Trp Ala
                65                  70                  75
Ala Gly Ile Trp Phe Thr Gly Arg Leu Gly Leu Arg Glu His Lys
                80                  85                  90
Ser Leu Ala Gln Gly Asp Ser Val Cys Pro Cys Glu Ser Glu Leu
```

```
                95                  100                 105
Gly Asp Phe Gln Val Tyr Gly Leu Val Ser Thr Glu Gly Val Val
                110                 115                 120
Ser Cys Phe Gly Glu Lys Thr Pro Gln His Pro Gly Pro Pro Ala
                125                 130                 135
Ser Leu Ser Leu Ala Asn Arg Cys His Asn Val Val Thr Ala Val
                140                 145                 150
Gly Ala Trp Pro Ala His Gly Ser Ile Leu Gly Asn Val Pro Glu
                155                 160                 165
Ala Pro Val Gly Ala Asp Val Leu Gly Ala Gly Gly Cys Asp Trp
                170                 175                 180
Ala Asp Lys Glu Ala Leu Ala Pro Gly Gln Arg Ala Lys Val His
                185                 190                 195
Ile Leu Leu Glu Ser Ser Gly Gln Ser Asp Pro Ser Tyr Ala Val
                200                 205                 210
Leu Pro Asp Ser Trp Ala Ala Thr Glu Gly Phe Pro Thr Tyr Arg
                215                 220                 225
Ser Gln Val Ser Ser Pro Arg Ile Pro Gly Ser Ser Ile Trp Leu
                230                 235                 240
Gly Ser Gly Ser Gly Trp Pro Ile Leu Gly Glu Leu Arg Glu Cys
                245                 250                 255
Asp Gln Met Phe Ser Cys Met Leu Pro Thr Gly Cys Ala Ser Phe
                260                 265                 270
Gln Asp Pro Gly Arg Tyr Gly Asp Tyr Asp Leu Pro Met Asp Glu
                275                 280                 285
Asp Glu Asp Met Thr Lys Thr Arg Thr Phe Phe Ala Ala Lys Ile
                290                 295                 300
Val Ile Gly Ile Ala Leu Ala Gly Ile Met Leu Val Cys Gly Ile
                305                 310                 315
Gly Asn Phe Val Phe Ile Ala Ala Leu Thr Arg Tyr Lys Lys Leu
                320                 325                 330
Arg Asn Leu Thr Asn Leu Leu Ile Ala Asn Leu Ala Ile Ser Asp
                335                 340                 345
Phe Leu Val Ala Ile Ile Cys Cys Pro Phe Glu Met Asp Tyr Tyr
                350                 355                 360
Val Val Arg Gln Leu Ser Trp Glu His Gly His Val Leu Cys Ala
                365                 370                 375
Ser Val Asn Tyr Leu Arg Thr Val Ser Leu Tyr Val Ser Thr Asn
                380                 385                 390
Ala Leu Leu Ala Ile Ala Ile Asp Arg Tyr Leu Ala Ile Val His
                395                 400                 405
Pro Leu Lys Pro Arg Met Asn Tyr Gln Thr Ala Ser Phe Leu Ile
                410                 415                 420
Ala Leu Val Trp Met Val Ser Ile Leu Ile Ala Ile Pro Ser Ala
                425                 430                 435
Tyr Phe Ala Thr Glu Thr Val Leu Phe Ile Val Lys Ser Gln Glu
                440                 445                 450
Lys Ile Phe Cys Gly Gln Ile Trp Pro Val Asp Gln Gln Leu Tyr
                455                 460                 465
Tyr Lys Ser Tyr Phe Leu Phe Ile Phe Gly Val Glu Phe Val Gly
                470                 475                 480
Pro Val Val Thr Met Thr Leu Cys Tyr Ala Arg Ile Ser Arg Glu
                485                 490                 495
Leu Trp Phe Lys Ala Val Pro Gly Phe Gln Thr Glu Gln Ile Arg
                500                 505                 510
Lys Arg Leu Arg Cys Arg Arg Lys Thr Val Leu Val Leu Met Cys
                515                 520                 525
Ile Leu Thr Ala Tyr Val Leu Cys Trp Ala Pro Phe Tyr Gly Phe
                530                 535                 540
Thr Ile Val Arg Asp Phe Phe Pro Thr Val Phe Val Lys Glu Lys
                545                 550                 555
His Tyr Leu Thr Ala Phe Tyr Val Val Glu Cys Ile Ala Met Ser
                560                 565                 570
```

```
Asn Ser Met Ile Asn Thr Val Cys Phe Val Thr Val Lys Asn Asn
                575                 580                 585
Thr Met Lys Tyr Phe Lys Lys Met Met Leu Leu His Trp Arg Pro
                590                 595                 600
Ser Gln Arg Gly Ser Lys Ser Ser Ala Asp Leu Asp Leu Arg Thr
                605                 610                 615
Asn Gly Val Pro Thr Thr Glu Glu Val Asp Cys Ile Arg Leu Lys
                620                 625                 630


<210> 13
<211> 695
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482007CD1

<400> 13
Met Lys Met Lys Ser Gln Ala Thr Met Ile Cys Cys Leu Val Phe
  1               5                  10                  15
Phe Leu Ser Thr Glu Cys Ser His Tyr Arg Ser Lys Ile His Leu
                20                  25                  30
Lys Ala Gly Asp Lys Leu Gln Ser Pro Glu Gly Lys Pro Lys Thr
                35                  40                  45
Gly Arg Ile Gln Glu Lys Cys Glu Gly Pro Cys Ile Ser Ser Ser
                50                  55                  60
Asn Cys Ser Gln Pro Cys Ala Lys Asp Phe His Gly Glu Ile Gly
                65                  70                  75
Phe Thr Cys Asn Gln Lys Lys Trp Gln Lys Ser Ala Glu Thr Cys
                80                  85                  90
Thr Ser Leu Ser Val Glu Lys Leu Phe Lys Asp Ser Thr Gly Ala
                95                 100                 105
Ser Arg Leu Ser Val Ala Ala Pro Ser Ile Pro Leu His Ile Leu
               110                 115                 120
Asp Phe Arg Ala Pro Glu Thr Ile Glu Ser Val Ala Gln Gly Ile
               125                 130                 135
Arg Lys Asn Cys Pro Phe Asp Tyr Ala Cys Ile Thr Asp Met Val
               140                 145                 150
Lys Ser Ser Glu Thr Thr Ser Gly Asn Ile Ala Phe Ile Val Glu
               155                 160                 165
Leu Leu Lys Asn Ile Ser Thr Asp Leu Ser Asp Asn Val Thr Arg
               170                 175                 180
Glu Lys Met Lys Ser Tyr Ser Glu Val Ala Asn His Ile Leu Asp
               185                 190                 195
Thr Ala Ala Ile Ser Asn Trp Ala Phe Ile Pro Asn Lys Asn Ala
               200                 205                 210
Ser Ser Asp Leu Leu Gln Ser Val Asn Leu Phe Ala Arg Gln Leu
               215                 220                 225
His Ile His Asn Asn Ser Glu Asn Ile Val Asn Glu Leu Phe Ile
               230                 235                 240
Gln Thr Lys Gly Phe His Ile Asn His Asn Thr Ser Glu Lys Ser
               245                 250                 255
Leu Asn Phe Ser Met Ser Met Asn Asn Thr Thr Glu Asp Ile Leu
               260                 265                 270
Gly Met Val Gln Ile Pro Arg Gln Glu Leu Arg Lys Leu Trp Pro
               275                 280                 285
Asn Ala Ser Gln Ala Ile Ser Ile Ala Phe Pro Thr Leu Gly Ala
               290                 295                 300
Ile Leu Arg Glu Ala His Leu Gln Asn Val Ser Leu Pro Arg Gln
               305                 310                 315
Val Asn Gly Leu Val Leu Ser Val Val Leu Pro Glu Arg Leu Gln
```

```
                320                 325                 330
Glu Ile Ile Leu Thr Phe Glu Lys Ile Asn Lys Thr Arg Asn Ala
                335                 340                 345
Arg Ala Gln Cys Val Gly Trp His Ser Lys Lys Arg Arg Trp Asp
                350                 355                 360
Glu Lys Ala Cys Gln Met Met Leu Asp Ile Arg Asn Glu Val Lys
                365                 370                 375
Cys Arg Cys Asn Tyr Thr Ser Val Val Met Ser Phe Ser Ile Leu
                380                 385                 390
Met Ser Ser Lys Ser Met Thr Asp Lys Val Leu Asp Tyr Ile Thr
                395                 400                 405
Cys Ile Gly Leu Ser Val Ser Ile Leu Ser Leu Val Leu Cys Leu
                410                 415                 420
Ile Ile Glu Ala Thr Val Trp Ser Arg Val Val Val Thr Glu Ile
                425                 430                 435
Ser Tyr Met Arg His Val Cys Ile Val Asn Ile Ala Val Ser Leu
                440                 445                 450
Leu Thr Ala Asn Val Trp Phe Ile Ile Gly Ser His Phe Asn Ile
                455                 460                 465
Lys Ala Gln Asp Tyr Asn Met Cys Val Ala Val Thr Phe Phe Ser
                470                 475                 480
His Phe Phe Tyr Leu Ser Leu Phe Phe Trp Ile Leu Phe Lys Ala
                485                 490                 495
Leu Leu Ile Ile Tyr Gly Ile Leu Val Ile Phe Arg Arg Met Met
                500                 505                 510
Lys Ser Arg Met Met Val Ile Gly Phe Ala Ile Gly Tyr Gly Cys
                515                 520                 525
Pro Leu Ile Ile Ala Val Thr Thr Val Ala Ile Thr Gly Pro Val
                530                 535                 540
Lys Gly Tyr Met Arg Pro Glu Ala Cys Trp Leu Asn Trp Asp Asn
                545                 550                 555
Thr Lys Ala Leu Leu Ala Phe Ala Ile Pro Ala Phe Val Ile Val
                560                 565                 570
Ala Val Asn Leu Ile Val Val Leu Val Val Ala Val Asn Thr Gln
                575                 580                 585
Arg Pro Ser Ile Gly Ser Ser Lys Ser Gln Asp Val Val Ile Ile
                590                 595                 600
Met Arg Ile Ser Lys Asn Val Ala Ile Leu Thr Pro Leu Leu Gly
                605                 610                 615
Leu Thr Trp Gly Phe Gly Ile Ala Thr Leu Ile Glu Gly Thr Ser
                620                 625                 630
Leu Thr Phe His Ile Ile Phe Ala Leu Leu Asn Ala Phe Gln Gly
                635                 640                 645
Phe Phe Ile Leu Leu Phe Gly Thr Ile Met Asp His Lys Ile Arg
                650                 655                 660
Asp Ala Leu Arg Met Arg Met Ser Ser Leu Lys Gly Lys Ser Arg
                665                 670                 675
Ala Ala Glu Asn Ala Ser Leu Gly Pro Thr Asn Gly Ser Lys Leu
                680                 685                 690
Met Asn Arg Gln Gly
                695

<210> 14
<211> 633
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6769042CD1

<400> 14
Met Tyr Phe Thr Ala Ala Ile Gly Lys His Ala Leu Leu Ser Ser
```

```
  1               5                  10                  15
Thr Leu Pro Ser Leu Phe Met Thr Ser Thr Ala Ser Pro Val Met
             20                  25                  30
Pro Thr Asp Ala Tyr His Pro Ile Ile Thr Asn Leu Thr Glu Glu
             35                  40                  45
Arg Lys Thr Phe Gln Ser Pro Gly Val Ile Leu Ser Tyr Leu Gln
             50                  55                  60
Asn Val Ser Leu Ser Leu Pro Ser Lys Ser Leu Ser Glu Gln Thr
             65                  70                  75
Ala Leu Asn Leu Thr Lys Thr Phe Leu Lys Ala Val Gly Glu Ile
             80                  85                  90
Leu Leu Leu Pro Gly Trp Ile Ala Leu Ser Glu Asp Ser Ala Val
             95                 100                 105
Val Leu Ser Leu Ile Asp Thr Ile Asp Thr Val Met Gly His Val
            110                 115                 120
Ser Ser Asn Leu His Gly Ser Thr Pro Gln Val Thr Val Glu Gly
            125                 130                 135
Ser Ser Ala Met Ala Glu Phe Ser Val Ala Lys Ile Leu Pro Lys
            140                 145                 150
Thr Val Asn Ser Ser His Tyr Arg Phe Pro Ala His Gly Gln Ser
            155                 160                 165
Phe Ile Gln Ile Pro His Glu Ala Phe His Arg His Ala Trp Ser
            170                 175                 180
Thr Val Val Gly Leu Leu Tyr His Ser Met His Tyr Tyr Leu Asn
            185                 190                 195
Asn Ile Trp Pro Ala His Thr Lys Ile Ala Glu Ala Met His His
            200                 205                 210
Gln Asp Cys Leu Leu Phe Ala Thr Ser His Leu Ile Ser Leu Glu
            215                 220                 225
Val Ser Pro Pro Pro Thr Leu Ser Gln Asn Leu Ser Gly Ser Pro
            230                 235                 240
Leu Ile Thr Val His Leu Lys His Arg Leu Thr Arg Lys Gln His
            245                 250                 255
Ser Glu Ala Thr Asn Ser Ser Asn Arg Val Phe Val Tyr Cys Ala
            260                 265                 270
Phe Leu Asp Phe Ser Ser Gly Glu Gly Val Trp Ser Asn His Gly
            275                 280                 285
Cys Ala Leu Thr Arg Gly Asn Leu Thr Tyr Ser Val Cys Arg Cys
            290                 295                 300
Thr His Leu Thr Asn Phe Ala Ile Leu Met Gln Val Val Pro Leu
            305                 310                 315
Glu Leu Ala Arg Gly His Gln Val Ala Leu Ser Ser Ile Ser Tyr
            320                 325                 330
Val Gly Cys Ser Leu Ser Val Leu Cys Leu Val Ala Thr Leu Val
            335                 340                 345
Thr Phe Ala Val Leu Ser Ser Val Ser Thr Ile Arg Asn Gln Arg
            350                 355                 360
Tyr His Ile His Ala Asn Leu Ser Phe Ala Val Leu Val Ala Gln
            365                 370                 375
Val Leu Leu Leu Ile Ser Phe Arg Leu Glu Pro Gly Thr Thr Pro
            380                 385                 390
Cys Gln Val Met Ala Val Leu Leu His Tyr Phe Phe Leu Ser Ala
            395                 400                 405
Phe Ala Trp Met Leu Val Glu Gly Leu His Leu Tyr Ser Met Val
            410                 415                 420
Ile Lys Val Phe Gly Ser Glu Asp Ser Lys His Arg Tyr Tyr Tyr
            425                 430                 435
Gly Met Gly Trp Gly Phe Pro Leu Leu Ile Cys Ile Ile Ser Leu
            440                 445                 450
Ser Phe Ala Met Asp Ser Tyr Gly Thr Ser Asn Asn Cys Trp Leu
            455                 460                 465
Ser Leu Ala Ser Gly Ala Ile Trp Ala Phe Val Ala Pro Ala Leu
            470                 475                 480
```

```
Thr Cys Ser Ser His Ile Met Val Val Ser Val Phe Tyr Gly Ala
                245                 250                 255
Ala Phe Tyr Thr Asn Val Leu Pro His Ser Tyr His Thr Pro Glu
                260                 265                 270
Lys Asp Lys Val Val Ser Ala Phe Tyr Thr Ile Leu Thr Pro Met
                275                 280                 285
Leu Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Ala Ala
                290                 295                 300
Ala Leu Arg Lys Val Leu Gly Arg Cys Gly Ser Ser Gln Ser Ile
                305                 310                 315
Arg Val Ala Thr Val Ile Arg Lys Gly
                320
```

<210> 17
<211> 315
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55036418CD1

<400> 17

```
Met Glu Thr Trp Val Asn Gln Ser Tyr Thr Asp Gly Phe Phe Leu
  1               5                  10                  15
Leu Gly Ile Phe Ser His Ser Thr Ala Asp Leu Val Leu Phe Ser
                20                  25                  30
Val Val Met Ala Val Phe Thr Val Ala Leu Cys Gly Asn Val Leu
                35                  40                  45
Leu Ile Phe Leu Ile Tyr Met Asp Pro His Leu His Thr Pro Met
                50                  55                  60
Tyr Phe Phe Leu Ser Gln Leu Ser Leu Met Asp Leu Met Leu Val
                65                  70                  75
Cys Thr Asn Val Pro Lys Met Ala Ala Asn Phe Leu Ser Gly Arg
                80                  85                  90
Lys Ser Ile Ser Phe Val Gly Cys Gly Ile Gln Ile Gly Leu Phe
                95                 100                 105
Val Cys Leu Val Gly Ser Glu Gly Leu Leu Leu Gly Leu Met Ala
                110                 115                 120
Tyr Asp Arg Tyr Val Ala Ile Ser His Pro Leu His Tyr Pro Ile
                125                 130                 135
Leu Met Asn Gln Arg Val Cys Leu Gln Ile Thr Gly Ser Ser Trp
                140                 145                 150
Ala Phe Gly Ile Ile Asp Gly Leu Ile Gln Met Val Val Val Met
                155                 160                 165
Asn Phe Pro Tyr Cys Gly Leu Arg Lys Val Asn His Phe Phe Cys
                170                 175                 180
Glu Met Leu Ser Leu Leu Lys Leu Ala Cys Val Asp Thr Ser Leu
                185                 190                 195
Phe Glu Lys Val Ile Phe Ala Cys Cys Val Phe Met Leu Leu Phe
                200                 205                 210
Pro Phe Ser Ile Ile Val Ala Ser Tyr Ala His Ile Leu Gly Thr
                215                 220                 225
Val Leu Gln Met His Ser Ala Gln Ala Trp Lys Lys Ala Leu Ala
                230                 235                 240
Thr Cys Ser Ser His Leu Thr Ala Val Thr Leu Phe Tyr Gly Ala
                245                 250                 255
Ala Met Phe Ile Tyr Leu Arg Pro Arg His Tyr Arg Ala Pro Ser
                260                 265                 270
His Asp Lys Val Ala Ser Ile Phe Tyr Thr Val Leu Thr Pro Met
                275                 280                 285
Leu Asn Pro Leu Ile Tyr Ser Leu Arg Asn Arg Glu Val Met Gly
                290                 295                 300
```

```
Ala Leu Arg Lys Gly Leu Asp Arg Cys Arg Ile Gly Ser Gln His
                305                 310                 315


<210> 18
<211> 324
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481701CD1

<400> 18
Met Glu Ser Pro Asn Gln Thr Thr Ile Gln Glu Phe Ile Phe Ser
  1               5                  10                  15
Ala Phe Pro Tyr Ser Trp Val Lys Ser Val Val Cys Phe Val Pro
                20                  25                  30
Leu Leu Phe Ile Tyr Ala Phe Ile Val Val Gly Asn Leu Val Ile
                35                  40                  45
Ile Thr Val Val Gln Leu Asn Thr His Leu His Thr Pro Met Tyr
                50                  55                  60
Thr Phe Ile Ser Ala Leu Ser Phe Leu Glu Ile Trp Tyr Thr Thr
                65                  70                  75
Ala Thr Ile Pro Lys Met Leu Ser Ser Leu Leu Ser Glu Arg Ser
                80                  85                  90
Ile Ser Phe Asn Gly Cys Leu Leu Gln Met Tyr Phe Phe His Ser
                95                 100                 105
Thr Gly Ile Cys Glu Val Cys Leu Leu Thr Val Met Ala Phe Asp
               110                 115                 120
His Tyr Leu Ala Ile Cys Ser Pro Leu His Tyr Pro Ser Ile Met
               125                 130                 135
Thr Pro Lys Leu Cys Thr Gln Leu Thr Leu Ser Cys Cys Val Cys
               140                 145                 150
Gly Phe Ile Thr Pro Val Pro Glu Ile Ala Trp Ile Ser Thr Leu
               155                 160                 165
Pro Phe Cys Gly Ser Asn His Leu Glu His Ile Phe Cys Asp Phe
               170                 175                 180
Leu Pro Val Leu Arg Leu Ala Cys Thr Asp Thr Arg Ala Ile Val
               185                 190                 195
Met Ile Gln Val Val Asp Val Ile His Ala Val Glu Ile Ile Thr
               200                 205                 210
Ala Val Met Leu Ile Phe Met Ser Tyr Asp Gly Ile Val Ala Val
               215                 220                 225
Ile Leu Arg Ile His Ser Ala Gly Gly Arg Arg Thr Ala Phe Ser
               230                 235                 240
Thr Cys Val Ser His Phe Ile Val Phe Ser Leu Phe Phe Gly Ser
               245                 250                 255
Val Thr Leu Met Tyr Leu Arg Phe Ser Ala Thr Tyr Ser Leu Phe
               260                 265                 270
Trp Asp Ile Ala Ile Ala Leu Ala Phe Ala Val Leu Ser Pro Phe
               275                 280                 285
Phe Asn Pro Ile Ile Tyr Ser Leu Arg Asn Lys Glu Ile Lys Glu
               290                 295                 300
Ala Ile Lys Lys His Ile Gly Gln Ala Lys Ile Phe Phe Ser Val
               305                 310                 315
Arg Pro Gly Thr Ser Ser Lys Ile Phe
               320

<210> 19
<211> 312
<212> PRT
<213> Homo sapiens
```

<220>
<221> misc_feature
<223> Incyte ID No: 7481774CD1

<400> 19
Met Glu Pro Trp Gln His Pro Thr His Phe Ile Leu Leu Gly Phe
1 5 10 15
Ser Asp Arg Pro His Leu Glu Arg Ile Leu Phe Val Val Ile Leu
20 25 30
Ile Ala Tyr Leu Leu Thr Leu Val Gly Asn Thr Thr Ile Ile Leu
35 40 45
Val Ser Arg Leu Asp Pro His Leu His Thr Pro Met Tyr Phe Phe
50 55 60
Leu Ala His Leu Ser Phe Leu Asp Leu Ser Phe Thr Thr Ser Ser
65 70 75
Ile Pro Gln Leu Leu Tyr Asn Leu Asn Gly Cys Asp Lys Thr Ile
80 85 90
Ser Tyr Met Gly Cys Ala Ile Gln Leu Phe Leu Phe Leu Gly Leu
95 100 105
Gly Gly Val Glu Cys Leu Leu Leu Ala Val Met Ala Tyr Asp Arg
110 115 120
Cys Val Ala Ile Cys Lys Pro Leu His Tyr Met Val Ile Met Asn
125 130 135
Pro Arg Leu Cys Arg Gly Leu Val Ser Val Thr Trp Gly Cys Gly
140 145 150
Val Ala Asn Ser Leu Ala Met Ser Pro Val Thr Leu Arg Leu Pro
155 160 165
Arg Cys Gly His His Glu Val Asp His Phe Leu Cys Glu Met Pro
170 175 180
Ala Leu Ile Arg Met Ala Cys Ile Ser Thr Val Ala Ile Asp Gly
185 190 195
Thr Val Phe Val Leu Ala Val Gly Val Val Leu Ser Pro Leu Val
200 205 210
Phe Ile Leu Leu Ser Tyr Ser Tyr Ile Val Arg Ala Val Leu Gln
215 220 225
Ile Arg Ser Ala Ser Gly Arg Gln Lys Ala Phe Gly Thr Cys Gly
230 235 240
Ser His Leu Thr Val Val Ser Leu Phe Tyr Gly Asn Ile Ile Tyr
245 250 255
Met Tyr Met Gln Pro Gly Ala Ser Ser Ser Gln Asp Gln Gly Lys
260 265 270
Phe Leu Thr Leu Phe Tyr Asn Ile Val Thr Pro Leu Leu Asn Pro
275 280 285
Leu Ile Tyr Thr Leu Arg Asn Arg Glu Val Lys Gly Ala Leu Gly
290 295 300
Arg Leu Leu Leu Gly Lys Arg Glu Leu Gly Lys Glu
305 310

<210> 20
<211> 1076
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474806CB1

<400> 20
caaagttgaa tgccgggttg gggcagaggc tgatgccgtg ctgaggtcat gatgttatgc 60
tgtccatttt gcttccttcc aggggaagca gaagcgggag ccgtcgtgga gctctgctcc 120
tggagggagc ctcccgggac atggagaagg tggacatgaa tacatcacag gaacaaggtc 180
tctgccagtt ctcagagaag tacaagcaag tctacctctc cctggcctac agtatcatct 240
ttatcctagg gctgccacta aatggcactg tcttgtggca ctcctggggc caaaccaagc 300

gctggagctg tgccaccacc tatctggtga acctgatggt ggccgacctg ctttatgtgc 360
tattgccctt cctcatcatc acctactcac tagatgacag gtggcccttc ggggagctgc 420
tctgcaagct ggtgcacttc ctgttctata tcaaccttta cggcagcatc ctgctgctga 480
cctgcatctc tgtgcaccag ttcctaggtg tgtggcaccc actgtgttcg ctgccctacc 540
ggacccgcag gcatgcctgg ctgggcacca gcaccacctg ggccctggtg gtcctccagc 600
tgctgcccac actggccttc tcccacacgg actacatcaa tggccagatg atctggtatg 660
acatgaccag ccaagagaat tttgatcggc tttttgccta cggcatagtt ctgacattgt 720
ctggctttct ttccccctcc ttggtcattt tggtgtgcta ttcactgatg gtcaggagcc 780
tgatcaagcc agaggagaac ctcatgagga caggcaacac agcccgagcc aggtccatcc 840
ggaccatcct actggtgtgt ggcctcttca ccctctgttt tgtgcccttc catatcactc 900
gctccttcta cctcaccatc tgctttctgc tttctcagga ctgccagctc ttgatggcac 960
ccagtgtggc ctacaagata tggaggcctc tggtgagtgt gagcagctgc ctcaacccag 1020
tcctgtactt tctttcaagg ggggcaaaaa tagagtcagg ctcctccaga aactga 1076

<210> 21
<211> 1102
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474840CB1

<400> 21
ggaggctgag gcaggagaat ggcatgaccc caggaggcag agcttgcagt gagatgagat 60
catgccactg tgctccagcc tgggcaacag agcgagactc tgtctcaaaa aaaaaaaaaa 120
acaaaaaaaa aaacctttt tcccaagcta ccattggact tttagccaac acctttttcc 180
ttttcttcaa catcttcata ttccttcagg atcagaaatc gaagccccat gacctcatca 240
gctgtaattc ggccttcatt catgtagtga tgttcctcac tgtggtggat gcttggcctc 300
cagatatgcc tgaatcactg cacttaggga atgagttcaa atttaagtcc ttgtcctaca 360
taaacagagt gaggatgggc ctatgtatct gtaacatctg tctcctgagt atacaccagg 420
ccaacaccat cagccccaac aacttctgtt tggcaaggct taaacagaaa ttcacaaata 480
acattatcat gtcatctttt ttttcttttt ttttttggtc catcaatttg tctttcagtt 540
ataacatagt attctttact gtggcttctt ctaatgtgac ccagaacagt ctacctaagg 600
gcagcaatac tgttcacttt ctccccatga agtccttcat gagaaaagta ttttttactc 660
tgacattatc cagggatgtc ttcattatag gaattacact gcattcaatt gcacacatgg 720
tgatccttgt gtccaggcat gagacgcaat ctcagcacct tcacagcatc agcatctctc 780
cacaagcctt cccagagaaa agggctgctc agaccatccc gctgttagtg agctactgtc 840
tggtcatgtg ctgggtggac ctcatcatct catcttcttc aaccctgctg tggacgtgta 900
acccagtctt cctgagtatg cagaaccttg tgggcgatgt ctatgccact gttgttctac 960
tggaacaaat cagctctgat aaaaatatag ttgacattct ccaaaatatg caaagtgcta 1020
taaagcttta acaagttggc gatggaaaac atttctaaaa aatagtcttc tcctatagtt 1080
caattgttca agtagccctg ga 1102

<210> 22
<211> 2529
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475092CB1

<400> 22
ggctccggtg tttcccgccg ttcatgcagc gaaaagagaa agcaaaatgc cctcaggagg 60
ctccagccgg ccgcgagccc tccacgcccg gcgggggcag cggaggcgga ggcgccgtcg 120
ctgcagcctc aggcgccgcg gtgccgggct ccgtgcagtt ggcgctgagc gtcctgcacg 180
ccctgctcta cgccgcgctg ttcgcctttg cctacctgca gctgtggcgg ctgctcctgt 240
accgcgagcg gcggctgagt taccagagcc tctgcctctt cctctgtctc ctgtgggcag 300
cgctcaggac caccctcttc tccgccgcct tctcgctcag cggctccctg cccttgctcc 360
ggccgcccgc tcacctgcac ttcttccccc actggctgct ctactgcttc ccctcctgtc 420
tccagttctc cacgctctgt ctcctcaacc tctacctggc ggaggttata tgtaaagtca 480
gatgtgccac tgaacttgac agacacaaaa ttctactgca tttgggcttt ataatggcaa 540

gcctgctctt tttagtggtg aacttgactt gcgcaatgct agttcatgga gatgtcccag 600
aaaatcagtt gaagtggact gtgtttgttc gagcattaat taatgatagc ctgtttattc 660
tttgtgccat ctctttagtg tgttacatat gcaaaattac aaaaatgtca tcagctaatg 720
tctacctcga atcaaagggt atgtctctgt gccagactgt cgtcgtgggc tctgtagtca 780
ttcttctgta ctcttccaga gcttgttata atttggtggt ggtcaccata tctcaggata 840
cattagaaag tccatttaat tatggctggg ataatctttc agataaggct catgtagaag 900
acataagtgg agaagagtat atagtatttg gaatggtcct ctttctgtgg gaacatgtgc 960
cagcatggtc ggtggtactg tttttccggg cacagagatt aaaccagaat ttggcacctg 1020
ctggcatgat aaatagtcac agttatagtt ccagagctta ctttttcgac aatccaagac 1080
gatatgatag tgatgatgac ctgccaagac tgggaagttc aagagaagga agtttaccaa 1140
attcgcaaag tttgggctgg tatggcacca tgactgggtg tggcagcagc agttacacag 1200
tcactcccca cctgaatgga cctatgacag atactgctcc tttgctcttt acttgtagta 1260
atttagattt gaacaatcat catagcttat atgtgacacc acaaaactga cagcatcacc 1320
aagtcatgat tcttgagttg tttttcataa atgtgtatat tcaatgtgtt taaattccat 1380
ctacataaac attccattat ctgttgcaac tgaaaacaaa atctggaagt gtggctgtgt 1440
ttggtaaata acacagctat tatttttgac ctcttcatag taaaatgaag taaaatggaa 1500
agtttggagt aggagaaaag agagattaga tcttaaggca cttgatggcc tccaaaaatc 1560
ctgactttgg aacatcaaat gcatatgtgc acttttatct ttgttctgag tcactgcagt 1620
ccccaaagtc atatgccaat gttcacactg aaatactgta ttgtacacca aactggaagg 1680
caattttcct atgaaaatca aagccggtat attcattggt atgctctata cagatatctt 1740
aataaaaatt ttatagtgtg aacagtgcac agagttaagg cataaaaatg tatcattctt 1800
tataaaaatc tactgaaaat gtgtaatcat tgaagacagt tcttttaagc atgattttaa 1860
aatagcaact gaaattcaat cattttaaac aaatgatggt agtaatccat tagttatggc 1920
cagcagtgtt ctttggagag ccacaataat ttcaagagga aaatatacca gtgaaaattg 1980
tgtggctatt ttgagtagaa ttggtcagtt gattattttg tgtaattgag atatatgtag 2040
tagtttaagc atgattcttg aagaaagcaa tagtgacttt tgcataggga gattttggta 2100
gaaacttctt gggactaaac aagtttagag atgcatttaa gaattattca caaaatgtgt 2160
aattctaaat taaaacataa atatattttc aaaagcattt gatttctctg aagcatgata 2220
tagctggtct tacctagtga atcaggattg tcctcaggta aatgaaatca tgatacatta 2280
ttgcagtgaa ctcaagtgca atactttgta agacatataa ttcctatgat tttcacattt 2340
ttatatctta tatatgggaa aagccaaatt aaattgaatt cagattaatt ccagcattag 2400
actaaatgag caaacttaag taaatgtaca aactaggtaa gtataaaacc acaggttaac 2460
aatattggag tacttttaga attacattaa aactgtctta aatgtcctat cccaaatcta 2520
aaaaaaaaa 2529

<210> 23
<211> 1847
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7341260CB1

<400> 23
gggggaggca ggcctcccag ttgctgcagt ttggaatatg tcaggtccca ccctcccaga 60
ggcggggcca gggctgagtc ctgccagcct catttctcta tccctctgag aacccagacg 120
ggcagagcct gggtaggaga gcctggcccc gctgtcccca ctgggtggag acaccatgca 180
cttggtccac ttgtgctctt cagccaggac accagacatg gtccaaaccg ctgcagggct 240
ggctgcagca actccctgac actcaggaag gcccaggctg ggcaggcaat acctgctccc 300
aacagccatg catgccggct gccgctccag gactcccctg tccccaggac caagatgacg 360
cccaacagca ctggcgaggt gcccagcccc attcccaagg gggctttggg gctctccctg 420
gccctggcaa gcctcatcat caccgcgaac ctgctcctag ccctgggcat cgcctgggac 480
cgccgcctgc gcagcccacc tgctggctgc ttcttcctga gcctactgct ggctgggctg 540
ctcacgggtc tggcattgcc cacattgcca gggctgtgga accagagtcg ccggggttac 600
tggtcctgcc tcctcgtcta cttggctccc aacttctcct tcctctccct gcttgccaac 660
ctcttgctgg tgcacgggga gcgctacatg gcagtcctga ggccactcca gccccctggg 720
agcattcggc tggccctgct cctcacctgg gctggtcccc tgctctttgc cagtctgccc 780
gctctggggt ggaaccactg gacccctggt gccaactgca gctcccaggc tatcttccca 840
gccccctacc tgtacctcga agtctatggg ctcctgctgc ccgccgtggg tgctgctgcc 900
ttcctctctg tccgcgtgct ggccactgcc caccgccagc tgcaggacat ctgccggctg 960
gagcgggcag tgtgccgcga tgagccctcc gccctggccc gggcccttac ctggaggcag 1020
gcaagggcac aggctggagc catgctgctc ttcgggctgt gctgggggcc ctacgtggcc 1080

acactgctcc tctcagtcct ggcctatgag cagcgcccgc cactggggcc tgggacactg 1140
ttgtccctcc tctccctagg aagtgccagt gcagcggcag tgcccgtagc catggggctg 1200
ggcgatcagc gctacacagc cccctggagg gcagccgccc aaaggtgcct gcaggggctg 1260
tggggaagag cctcccggga cagtcccggc cccagcattg cctaccaccc aagcagccaa 1320
agcagtgtcg acctggactt gaactaaagg aagggcctct gctgactcct accagagcat 1380
ccgtccagct cagccatcca gcctgtctct actgggcccc acttctctgg atcagagacc 1440
ctgcctctgt ttgaccccgc actgactgaa taaagctcct ctggccgtta aaaaaaaaaa 1500
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaac aaacaaacag 1560
aaaaaaaaaa aaaagaacac acaaagaaca cagaacaaac caagcagcac accacacaca 1620
aaaacaatgc acacacagaa caagacacaa tcagagacag agagcacaca gcacggaccc 1680
cagccacgcc cccagcactg accaccacga cccgacacag aaacgaacac tgaagactca 1740
acgcacaaaa cgacaaccag accacaagcc aaccgcctca cgccccagca acgaacacac 1800
atacaaaacc aaaccgagac aacccacata cagccaaaca aaccaca 1847

<210> 24
<211> 2031
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473911CB1

<400> 24
atgaataaaa acaacaaacc ttccagtttc atagccataa gaaatgctgc tttctctgaa 60
gtcggcattg ggatctctgc caatgccatg ctccttctct tccacatcct cacgtgcctt 120
ctcaagcaca ggaccaagcc cgctgacctg atcgtttgtc atgtggctct aatccatatc 180
atattgctgc tacccacaga gttcatagct acagatattt ttgggtctca ggattcagag 240
gatgacatca aacataagtc agttatctac aggcgtaaca ggcagtccca gcattttcac 300
agcaccaacc tttctccaaa agcaccccca gaaaaaatgg ccacgcagac cattcttctg 360
ctcgtgagtt gctttgtgat tgtgtatgtt ttggactgtg ttgtcgcctc ctgctcagga 420
ctggtgtgga acagtgatcc agtccgtcat cgagtccaga tgctggtgga caatggctat 480
gccaccatca gtccttcagt gctacccagg ctgactgccc caaacgagtg gagagccagt 540
gtgtacctga atgacagctt gaacaaatgc agcaacggac ggctgctctg tgtagacagg 600
gggcttgatg aggggccccg gtccgtccca aagtgctctg agtcagagac cgacgaggat 660
tacatcgtcc tcagggctcc gctgagggag gacgaaccca aggacggggg cagtgtgggg 720
aatgcagccc tggtgtctcc cgaggcctct gcagaagagg aagaggagcg tgaggaggga 780
ggcgaggcat gtggcctgga gaggacagga gctggtgggg agcaggttga ccttggtgaa 840
ctacctgacc atgaggagaa aagcaaccag aaagtggcag ctgccaccct ggaggaccgc 900
acacaggatg agcctgctga ggagagctgc cagatcgtcc ttttccagaa caactgcatg 960
gacaactttg tgacttccct cacaggaagc ccctacgagt tcttcccaac caagagcacc 1020
tcttttgca gggagagctg ttctcctttt tctgagtcag tgaaaagctt agaatcagag 1080
caggcaccaa agttggggct gtgtgcggag gaggaccccg tggttggggc tttgtgtggc 1140
cagcatggac ccttgcaaga tggagtggcg gagggtccca cagcccctga tgtggtggtc 1200
ctgccgaagg aggaggagaa ggaggaggtc attgtggatg acatgctggc caacccctat 1260
gtgatgggag atgaggggga ggaggaggag gaggagttcg tggatgacac actggccaac 1320
ccctatgtga tgggagtggg cctgccagga agaggagggg aggaggagga ggaggaggag 1380
gtcgtggatg acacgctggc cagcctctat aagatgggag aagaacatcg acacaagggc 1440
ctggccccac tctgggaagg tggccagaaa ccgtcccaga aactgccccc aaagaaacca 1500
gatctgaggc aggttcctca gcccctggca tcggaggtgc cgcagaggag gcaggaaaga 1560
gctgttgtca ctgaagggag gcccctggaa gccagcaggg ccttgccagc aaagcccagg 1620
gccttcactt tataccctcg gtcgttctcc gtggaaggcc aagagattcc tgtttccatc 1680
tctgtgtact gggagccaga agggtcgggg ttagatgacc acaggataaa gaggaaagag 1740
gaacatctct ctgtcgtgtc tgggagtttc tcccagagaa accaccttcc atccagcggc 1800
acctccacgc cttcttccat ggtcgacatc ccacctcctt tcgacctggc ctgcatcacc 1860
aagaagccca tcacaaagag ctctccctct ctcctgatcg acagcgactc cccggacaag 1920
tacaagaaga agaagtcatc ctttaagcgg ttcctggcgc tgatgtttaa caagatggag 1980
aggccaggca cgatggctca tgcctgtcat cccagcactt tgggaagctg a 2031

<210> 25
<211> 1130
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474767CB1

<400> 25
ggggcgcgct catggagcac acgcacgccc acctcgcagc caacagctcg ctgtcttggt 60
ggtcccccgg ctcggcctgc ggcttgggtt tcgtgcccgt ggtctactac agcctcttgc 120
tgtgcctcgg tttaccagca aatatcttga cagtgatcat cctctcccag ctggtggcaa 180
gaagacagaa gtcctcctac aactatctct tggcactcgc tgctgccgac atcttggtcc 240
tctttttcat agtgtttgtg gacttcctgt tggaagattt catcttgaac atgcagatgc 300
ctcaggtccc cgacaagatc atagaagtgc tggaattctc atccatccac acctccatat 360
ggattactgt accgttaacc attgacaggt atatcgctgt ctgccacccg ctcaagtacc 420
acacggtctc atacccagcc cgcacccgga aagtcattgt aagtgtttac atcacctgct 480
tcctgaccag catcccctat tactggtggc ccaacatctg gactgaagac tacatcagca 540
cctctgtgca tcacgtcctc atctggatcc actgcttcac cgtctacctg gtgccctgct 600
ccatcttctt catcttgaac tcaatcattg tgtacaagct caggaggaag agcaattttc 660
gtctccgtgg ctactccacg gggaagacca ccgccatctt gttcaccatt acctccatct 720
ttgccacact ttgggccccc cgcatcatca tgattcttta ccacctctat ggggcgccca 780
tccagaaccg ctggctggta cacatcatgt ccgacattgc caacatgcta gcccttctga 840
acacagccat caacttcttc ctctactgct tcatcagcaa gcggttccgc accatggcag 900
ccgccacgct caaggctttc ttcaagtgcc agaagcaacc tgtacagttc tacaccaatc 960
ataacttttc cataacaagt agcccctgga tctcgccggc aaactcacac tgcatcaaga 1020
tgctggtgta ccagtatgac aaaaatggaa aacctataaa aagtcgtaat gacagcaaaa 1080
gctcctacca gtttgaagat gccattggag cttgtgtcat catcctgtga 1130

<210> 26
<211> 1202
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475815CB1

<400> 26
caggttctgc aatacaattg gaaacaactc attgctcctg cctatgaaga agtagaacta 60
tggtaaaaat aagaaaatgc cttcccaata atagggagtt gtactatgga agtaatatgg 120
aggcccagca atgggaatca tagtccagcc atgatggcat agtcatggga gaagagagag 180
ggtggggatg gcttccagga gggtgtgaag agggggtcca gtactgaagg gggaaaagat 240
gcatcagaat taattaatgt attttgatga tggcaatagt gttggttgag attggtgaag 300
gtagtaatat ttgtgatatt tttgttgctt ttctccctag acattaacta tgtgcttatt 360
ttccccataa gatgaataaa aacaacaaac cttccagttt catagccata agaaatgctg 420
ctttctctga agtcggcatt gggatctctg ccaatgccat gctccttctc ttccacatcc 480
tcacgtgcct tctcaagcac aggaccaagc ccgctgacct gatcgtttgt catgtggctc 540
taatccatat catattgctg ctacccacag agttcatagc tacagatatt tttgggtctc 600
aggattcaga ggatgacatc aaacataagt cagttatcta caggtacagg ttgatgagag 660
gcctctccat ttccaccacc tgcctgctga gtatcctccc ggccatcacc tgcagcccca 720
gaagctcctg tttggcagtg ttcaaagatt ctcacatcac caaccacgtt gctttctctt 780
ccgtcttcca catatccatt agtgacagct tcttagtctc cactcttccc atcaaaaatc 840
tggcctcaaa tagccttaca tttgtcactc aatcctgctc tgctgggatc ggctcacggc 900
ccccctccag tggatacatg gtgattctct tgtccaggcg taacaggcag tcccagcatt 960
ttcacagcac caacctttct ccaaaagcac ccccagaaaa aatggccacg cagaccattc 1020
ttctgctcgt gagttgcttt gtgattgtgt atgttttgga ctgtgttgtc gcctcctgct 1080
caggactggt gtggaacagt gatccagtcc gtcatcgagt ccagatgctg gtggacaatg 1140
gctatgccac catcagtcct tcagtgctag tcagtactga aaaatgaatg atcaaagtct 1200
ga 1202

<210> 27
<211> 2079
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 60263275CB1

<400> 27
tacggcacag tagagagctt ccagggctgg ctggcgtggg atacccgtac cacagaaatg 60
cagggaccat tgcttcttcc aggcctctgc tttctgctga gcctctttgg agctgtgact 120
cagaaaacca aaaacattaa tgaatgtaca ccaccctata gtgtatattg tggatttaac 180
gctgtgtgtt acaatgtcga aggaagtttc tactgtcaat gtgtcccagg atatagactg 240
cattctggga atgaacaatt cagtaattcc aatgagaaca cctgtcagga caccacctcc 300
tcaaagacaa cccagggcag gaaagagctg caaaagattg tggacaaatt tgagtcactt 360
ctcaccaatc agactttatg gagaacagaa gggagacaag aaatctcatc cacagctacc 420
actattctcc gggatgtgga atcgaaagtt ctagaaactg ccttgaaaga tccagaacaa 480
aaagtcctga aaatccaaaa cgatagtgta gctattgaaa ctcaagcgat tacagacaat 540
tgctctgaag aaagaaagac attcaacttg aacgtccaaa tgaactcaat ggacatccgt 600
tgcagtgaca tcatccaggg agacacacaa ggtcccagtg ccattgcctt tatctcatat 660
tcttctcttg gaaacatcat aaatgcaact ttttttgaag agatggataa gaaagatcaa 720
gtgtatctga actctcaggt tgtgagtgct gctattggac ccaaaaggaa cgtgtctctc 780
tccaagtctg tgacgctgac tttccagcac gtgaagatga cccccagtac caaaaaggtc 840
ttctgtgtct actggaagag cacagggcag ggcagccagt ggtccaggga tggctgcttc 900
ctgatacacg tgaacaagag tcacaccatg tgtaattgca gtcacctgtc cagcttcgct 960
gtcctgatgg ccctgaccag ccaggaggag gatcccgtgc tgactgtcat cacctacgtg 1020
gggctgagcg tctctctgct gtgcctcctc ctggcggccc tcacttttct cctgtgtaaa 1080
gccatccaga acaccagcac ctcactgcat ctgcagctct cgctctgcct cttcctggcc 1140
cacctcctct tcctcgtggg gattgatcga actgaaccca aggtgctgtg ctccatcatc 1200
gccggtgctt tgcactatct ctacctggcc gccttcacct ggatgctgct ggagggtgtg 1260
cacctcttcc tcactgcacg gaacctgaca gtggtcaact actcaagcat caatagactc 1320
atgaagtgga tcatgttccc agtcggctat ggcgttcccg ctgtgactgt ggccatttct 1380
gcagcctcct ggcctcacct ttatggaact gctgatcgat gctggctcca cctggaccag 1440
ggattcatgt ggagtttcct tggcccagtc tgtgccattt tctctgcgaa tttagtattg 1500
tttatcttgg tcttttggat tttgaaaaga aaactttcct ccctcaatag tgaagtgtca 1560
accatccaga acacaaggat gctggctttc aaagcaacag ctcagctctt catcctgggc 1620
tgcacatggt gtctgggctt gctacaggtg ggtccagctg cccaggtcat ggcctacctc 1680
ttcaccatca tcaacagcct ccaaggcttc ttcatcttct tggtctactg cctcctcagc 1740
cagcaggtcc agaaacaata tcaaaagtgg tttagagaga tcgtaaaatc aaaatctgag 1800
tctgagacat acacactttc cagcaagatg ggtcctgact caaaacccag tgaggggat 1860
gtttttccag gacaagtgaa gagaaaatat taaaactaga atattcaact ccatatggaa 1920
aatcatatcc atggatctct ttggcattat gaagaatgaa gctaaggaaa agggaattca 1980
ttaaacatat catccttgga gaggaagtaa tcaaccttta cttcccaaac tgtttgttct 2040
ccacaatagg tctcaacaaa tgtgtggtaa attgcatta 2079

<210> 28
<211> 5324
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 60203310CB1

<400> 28
ggcggcagca gcagcagcaa gaggaggaga agcagcccca gctatgactg ccgcatgtta 60
atagctgccg ctgggtccct cggctgctgc tggagacaga gcctgactcc gaagttgtgc 120
aactgtggac tgggagagac atttgaaccc tcttttcttt tcgctccccc tttgcccccc 180
tggggtgtgt gaagcgagga acgtaaagga aggcgaacat ttggctctct ttttccttcc 240
cctttctccg tggctgtgta gcggaagaaa gggaagagag acttttttgtt gttgtttcct 300
tgactggggt ctccaccctc ctgctgcttt ctctgcgctt cgattctcgt tatttgccgc 360
gtgtggttgg gggtgtctgc acaggggccg gccggtcttt tgccccgggc tcaatggctg 420
gattgtggaa actgcacccg ccttcaggtt gttgagcaac tgatgggacg atctcaggga 480
ccggcgttta cgaaaggttt cagatttggg atattggtgt ttctgttttg gagaaattat 540
tctttttctt tttaatttga agaaaaatca tcagtcttgg aatacagaag agaaactaga 600
aatatacgta ttttgtttca catttgaaca gtcattcttg aggaatactc catacctgag 660
tagacagcca tgtggccatc gcagctacta attttcatga tgctcttagc tccaataatt 720
catgctttca gccgtgcccc aattccaatg gctgtggtcc gcagagagct atcctgtgag 780

agctatccta tagagcttcg ctgtccagga acagacgtca tcatgataga aagtgccaac 840
tatggcagga ctgatgacaa aatttgtgac tctgaccctg ctcagatgga gaatatccga 900
tgttatctgc cagatgccta taagattatg tctcaaagat gcaataacag aacccagtgt 960
gcagtggtgg caggtcctga tgtttttcca gacccgtgtc caggaaccta taaatacctt 1020
gaagtgcagt atgaatgtgt cccttacaaa gtggaacaaa aagtttttct ttgtcctgga 1080
ctactaaaag gagtatacca gagtgaacat ttgtttgagt ccgaccacca atctggggcg 1140
tggtgcaaag accctctgca ggcatctgac aagatttatt atatgccctg gactccctac 1200
agaactgata ccctgactga gtattcatcc aaggatgact tcattgctgg aagaccaact 1260
acaacctaca agctccctca tagggtggat ggcacaggat ttgtagtgta tgatggagct 1320
ttgttcttca acaaagagcg caccaggaac atagtaaagt ttgatttgcg gactaggata 1380
aagagtggag aggctatcat agcaaatgcc aattaccatg atacctcccc ttaccgatgg 1440
ggaggcaaat ctgacataga cctggcagta gatgagaatg ggctatgggt aatctatgca 1500
acagaacaaa acaatggtaa aattgtcatt agtcaattga acccttacac cctacggatc 1560
gaaggaacat gggatactgc atatgataaa aggtcagctt ccaatgcctt tatgatttgt 1620
ggaattctgt atgtggtcaa atctgtatat gaggatgatg acaatgaggc tactggaaat 1680
aagattgact acatttacaa cactgaccaa agcaaggata gtttggtgga tgtacccttt 1740
cctaattcat accagtacat tgcagctgtg gattacaacc ccagggacaa cctactttat 1800
gtatggaata actatcacgt cgtgaaatat tctttggatt ttggacctct ggatagtaga 1860
tcagggcagg cacatcatgg acaagtttca tacatttctc cgccaattca ccttgactct 1920
gagctagaaa gaccctctgt taaagatatc tctaccacag gacctcttgg catgggaagc 1980
actaccacca gtaccaccct tcggaccaca actttgagcc caggaaggag taccaccccg 2040
tcagtgtcag gaagaagaaa ccggagtact agtaccccat ctccagctgt cgaggtactt 2100
gatgacatga ccacacacct tccatcagca tcgtcccaaa tcccagctct cgaagagagc 2160
tgtgaggctg tggaagcccg agaaatcatg tggtttaaga ctcgtcaagg acagatagca 2220
aagcagccat gccctgcagg aactataggt gtatcaactt atctatgcct tgctcctgat 2280
ggaatttggg atccccaagg tccagatctc agcaactgtt cttctccttg ggtcaatcat 2340
ataacacaga agttgaaatc tggtgaaaca gctgccaaca ttgctagaga gctggctgaa 2400
cagacaagaa atcacttgaa tgctggggac atcacctact ctgtccgggc catggaccag 2460
ctggtaggcc tcctagatgt acagcttcgg aacttgaccc caggtggaaa agatagtgct 2520
gcccggagtt tgaacaagct tcagaaaaga gagcgctctt gcagagccta tgtccaggca 2580
atggtcgaga cagttaacaa cctccttcag ccacaagctt tgaatgcatg gagagacctg 2640
actacgagtg atcagctgcg tgcggccacc atgttgcttc atactgtgga ggaaagtgct 2700
tttgtgctgg ctgataacct tttgaagact gacattgtca gggagaacac agacaatatt 2760
aaattggaag ttgcaagact gagcacagaa ggaaacttag aagacctaaa atttccagaa 2820
aacatgggcc atggaagcac tatccagctg tctgcaaata ccttaaagca aaatggccga 2880
aatggagaga tcagagtggc ctttgtcctg tataacaact tgggtcctta tttatccacg 2940
gagaatgcca gtatgaagtt gggaacggaa gctttgtcca caaatcattc tgttattgtc 3000
aattcccctg ttattacggc agcaataaac aaagagttca gtaacaaggt ttatttggct 3060
gatcctgtgg tatttactgt taaacatatc aagcagtcag aggaaaattt caaccctaac 3120
tgttcatttt ggagctactc caagcgtaca atgacaggtt attggtcaac acaaggctgt 3180
cggctcctga caacaaataa gacacatact acatgctctt gtaaccacct aacaaatttt 3240
gcagtactga tggcacatgt ggaagttaag cacagtgatg cggtccatga cctccttctg 3300
gatgtgatca cgtgggttgg aattttgctg tcccttgttt gtctcctgat ttgcatcttc 3360
acattttgct ttttccgcgg gctccagagt gaccgtaaca ccatccacaa gaacctctgc 3420
atcagtctct ttgtagcaga gctgctcttc ctgattggga tcaaccgaac tgaccaaccg 3480
attgcctgtg ctgttttcgc tgccctgtta catttcttct tcttggctgc cttcacctgg 3540
atgttccttg agggggtgca gctttatatc atgctggtgg aggtttttga gagtgaacat 3600
tcacgtagga aatactttta tctggtcggc tatgggatgc ctgcactcat tgtggctgtg 3660
tcagctgcag tagactacag gagttatgga acagataaag tatgttggct ccgacttgac 3720
acctacttca tttggagttt tataggacca gcaactttga taattatgct taatgtaatc 3780
ttccttggga ttgctttata taaaatggtt catcatactg ctatactgaa acctgaatca 3840
ggctgtcttg ataacatcaa ctatgaggat aacagaccct tcatcaagtc atgggttata 3900
ggtgcaatag ctcttctctg cctattagga ttgacctggg cctttggact catgtatatt 3960
aatgaaagca cagtcatcat ggcctatctc ttcaccattt tcaattctct acagggaatg 4020
tttatattta ttttccattg tgtcctacag aagaaggtac gaaaagagta tgggaaatgc 4080
ctgcgaacac attgctgtag tggcaaaagt acagagagtt ccattggttc agggaaaaca 4140
tctggttctc gaactcctgg acgctactcc acaggctcac agagccgaat ccgtagaatg 4200
tggaatgaca cggttcgaaa gcagtcagag tcttccttta ttactggaga cataaacagt 4260
tcagcgtcac tcaacagaga ggggcttctg aacaatgcca gggatacaag tgtcatggat 4320
actctaccac tgaatggtaa ccatggcaat agttacagca ttgccagcgg cgaatacctg 4380
agcaactgtg tgcaaatcat agaccgtggc tataaccata acgagaccgc cctagagaaa 4440
aagattctga aggaactcac ttccaactat atcccttctt acctgaacaa ccatgagcgc 4500
tccagtgaac agaacaggaa tctgatgaac aagctggtga ataaccttgg cagtggaagg 4560

gaagatgatg ccattgtcct ggatgatgcc acctcgttta accacgagga gagtttgggc 4620
ctggaactca ttcatgagga atctgatgct cctttgctgc ccccaagagt atactccacc 4680
gagaaccacc agccacacca ttataccaga aggcggatcc cccaagacca cagtgagagc 4740
tttttccctt tgctaaccaa cgagcacaca gaagatctcc agtcacccca tagagactct 4800
ctctatacca gcatgccgac actggctggt gtggccgcca cagagagtgt taccaccagc 4860
acccagaccg aacccccacc ggccaaatgt ggtgatgccg aagatgttta ctacaaaagc 4920
atgccaaacc taggctccag aaaccacgtc catcagctgc atacttacta ccagctaggt 4980
cgcggcagca gtgatggatt tatagttcct ccaaacaaag atgggacccc tcccgaggga 5040
agttcaaaag gaccggctca tttggtcact agtctataga agatgacaca gaaattggaa 5100
ccaacaaaac tgctaacacc ttgttgactg ttctgagttg atataagcag tggtaataat 5160
gtgtgtactc ctaaatcttt atgctgtcct ctaaagacaa acacaaactc tcagactttt 5220
ttttttcaac tgggatttaa ggtcagccca ggggagaaag ataactgcta aaattccccc 5280
gtaccccatc ctttcttgtc ctttcccccc tcagatggag actt 5324

<210> 29
<211> 1962
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477349CB1

<400> 29
atggatccca gcgttgttag caatgagtat tatgatgttg cccatggagc aaaagatcca 60
gtggtcccca cttccctgca ggacatcact gctgtcctgg gtacagaagc atatactgag 120
gaagacaaat caatggtgtc ccatgcacag aaaagccagc attcttgtct cagccattcc 180
aggtggctga ggtctccaca ggtcacaggg ggaagctggg acctccgaat aaggccatcc 240
aaggactcca gcagtttccg ccaggctcag tgtctgcgta aggatcctgg ggcaaacaac 300
cacttggaga gccaaggggt gagaggtaca gctggcgatg ctgacaggga gctgcgggga 360
ccctcagaaa aagccacagc tggccagcca cgagtgaccc tgctgcccac gcccaacgtc 420
agcgggctga gccaggagtt tgaaagccac tggccagaga tcgcagagag gtccccgtgt 480
gtggctggcg tcatccctgt catctactac agtgtcctgc tgggcttggg gctgcctgtc 540
agcctcctga ccgcagtggc cctggcgcgc cttgccacca ggaccaggag gccctcctac 600
tactaccttc tggcgctcac agcctcggat atcatcatcc aggtggtcat cgtgttcgcg 660
ggcttcctcc tgcagggagc agtgctggcc cgccaggtgc cccaggctgt ggtgcgcacg 720
gccaacatcc tggagtttgc tgccaaccac gcctcagtct ggatcgccat cctgctcacg 780
gttgaccgct acactgccct gtgccacccc ctgcaccatc gggccgcctc gtccccaggc 840
cggacccgcc gggccattgc tgctgtcctg agtgctgccc tgttgaccgg catccccttc 900
tactggtggc tggacatgtg gagagacacc gactcaccca gaacactgga cgaggtcctc 960
aagtgggctc actgtctcac tgtctatttc atcccttgtg gcgtgttcct ggtcaccaac 1020
tcggccatca tccaccggct acggaggagg ggccggagtg ggctgcagcc ccgggtgggc 1080
aagagcacag ccatcctcct gggcatcacc acactgttca ccctcctgtg ggcgccccgg 1140
gtcttcgtca tgctctacca catgtacgtg gcccctgtcc accgggactg gagggtccac 1200
ctggccttgg atgtggccaa catggtggcc atgctccaca cggcagccaa cttcggcctc 1260
tactgctttg tcagcaagac tttccgggcc actgtccgac aggtcatcca cgatgcctac 1320
ctgccctgca ctttggcatc acagccagag ggcatggcgg cgaagcctgt gatggagcct 1380
ccgggactcc ccacaggggc agaagtgtag aggagggggc ccagctaggg agctcagggt 1440
ggctcatggc cacatgtact ggggcctttg aggttgtacc caaaacacgt ttatcaacag 1500
cttgctttcc ttgggtgggg gtggaggctc ctcctttggg tgtggctccc aggtagagag 1560
gaggacaact tagccagctc ttatgtttgc ttcaccagca atccctattt cctgggaaga 1620
tgaaagggca ctgccaggca caggctaata gcatcagtgc tgtgggcatt cctttgcggg 1680
gggcattttg cctggctcat cgtgaatgcc agattaatgt tggttgaatg gatagaaaaa 1740
cggcctctca ttttcgtaac tgaggcagga gaatcgcttg aacccaggag acggaggttg 1800
cagcgagctg agatcgcgcc atagaaacac catggaactc caacctgggc aacaagagtg 1860
aaacttcgac tcaaaaaaaa aagagaaaaa acacattagg taacagtttc tttttagcat 1920
ttgtgtaacc tttaataaaa taaagtgata atcaaaaaaa aa 1962

<210> 30
<211> 1558
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55002225CB1

<400> 30
attttattcg cgaaggcacc ccacgctcct agaaaagagc acgacgcacc cgatgctcgg 60
attggatgaa gtggcaaagc tttaatccct ggaaagtcca cgaacaatga atccatttca 120
tgcatcttgt tggaacacct ctgccgaact tttaaacaaa tcctggaata aagagtttgc 180
ttatcaaact gccagtgtgg tagatacagt catcctccct tccatgattg ggattatctg 240
ttcaacaggg ctggttggca acatcctcat tgtattcact ataataagat ccaggaaaaa 300
aacagtccct gacatctata tctgcaacct ggctgtggct gatttggtcc acatagttgg 360
aatgcctttt cttattcacc aatgggcccg agggggagag tgggtgtttg gggggcctct 420
ctgcaccatc atcacatccc tggatacttg taaccaattt gcctgtagtg ccatcatgac 480
tgtaatgagt gtggacaggt actttgccct cgtccaacca tttcgactga cacgttggag 540
aacaaggtac aagaccatcc ggatcaattt gggcctttgg gcagcttcct ttatcctggc 600
attgcctgtc tgggtctact cgaaggtcat caaatttaaa gacggtgttg agagttgtgc 660
ttttgatttg acatcccctg acgatgtact ctggtataca ctttatttga cgataacaac 720
ttttttttc cctctaccct tgattttggt gtgctatatt ttaattttat gctatacttg 780
ggagatgtat caacagaata aggatgccag atgctgcaat cccagtgtac caaaacagag 840
agtgatgaag ttgacaaaga tggtgctggt gctggtggta gtctttatcc tgagtgctgc 900
cccttatcat gtgatacaac tggtgaactt acagatggaa cagcccacac tggccttcta 960
tgtgggttat tacctctcca tctgtctcag ctatgccagc agcagcatta acccttttct 1020
ctacatcctg ctgagtggaa cgcctcaaat ccaaagaaga gcgactgaga aggaaatcaa 1080
caatatggga aacactctga aatcacactt ttaggaaagt acatggatca ccatgagtct 1140
agacatgatt gtctatctta ctggtattat tagaaagggc aggtgtaccg atatgtttat 1200
gcccattctt cttgtgtact tgtgactctt agcagcatgg aagagaagtg taaccatgca 1260
aatacaatga gcttaatatg ctaactttag caagatgtaa aatgttgatc tatattgtgg 1320
gtagggaatg ggatagtctg agatacccag gcttcatgat ggtgtatatt atttcagcat 1380
attataaact agtcactaat gaaaatggcc atccatgacc attgactcaa aactcaccaa 1440
ggaacctgac cttgccctcc acactgcggc ctcactgtaa cagtttcctc aaggttccta 1500
ggagggtatc accttagagt gaagtctaaa atttggctat tttttatcta ttaaaaat 1558

<210> 31
<211> 2304
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475686CB1

<400> 31
atgcggctgg gacctgtccc agcccgggcg cgcgccctct tgtcttgggt taggggggctg 60
gaaagccgag gaggggagtg gaccaaatgc attgttcagc tgggtcatct ccttgctacc 120
cagcatcccg cggcgcccac atgtggagtc gtttccagcg ccctggtcat gcactcaaca 180
gatgtctgtc tagcccccac tatgcaccag gcactggact gggcagcagg aatttggttt 240
acaggaagat taggactcag agagcataaa tcactggccc agggtgactc agtctgtcca 300
tgtgaaagtg aacttggtga tttccaagtc tatggcttgg tcagtacaga aggagtggtg 360
tcctgctttg gagagaagac cccgcagcat cctggccctc ctgcttcatt gtccctggcc 420
aacaggtgcc acaacgttgt gacagctgta ggagcctggc cagctcatgg gagcatcctt 480
ggaaatgttc cagaagcccc tgtgggagct gatgtgttgg gggctggagg atgtgactgg 540
gcagacaaag aggccctggc ccctgggcaa agggcaaagg tgcacattct tcttgagagt 600
tctggacagt ctgatccatc ctatgctgtc cttcctgaca gctgggcagc cacggagggt 660
ttcccaactt acagatctca ggtctcctct ccccgcatcc cgggtagttc catctggtta 720
ggcagtgggt ctggttggcc tatacttggg gaactcaggg aatgtgacca gatgttctcc 780
tgcatgttgc ccactggttg tgcctccttc caggatccag gacgttatgg tgattatgac 840
ctccctatgg atgaggatga ggacatgacc aagacccgga ccttcttcgc agccaagatc 900
gtcattggca ttgcactggc aggcatcatg ctggtctgcg gcatcggtaa ctttgtcttt 960
atcgctgccc tcacccgcta taagaagttg cgcaacctca ccaatctgct cattgccaac 1020
ctggccatct ccgacttcct ggtggccatc atctgctgcc ccttcgagat ggactactac 1080
gtggtacggc agctctcctg ggagcatggc cacgtgctct gtgcctccgt caactacctg 1140
cgcaccgtct ccctctacgt ctccaccaat gccttgctgg ccattgccat tgacaggtat 1200
ctcgccatcg ttcacccctt gaaaccacgg atgaattatc aaacggcctc cttcctgatc 1260

gccttggtct ggatggtgtc cattctcatt gccatcccat cggcttactt tgcaacagaa 1320
acggtcctct ttattgtcaa gagccaggag aagatcttct gtggccagat ctggcctgtg 1380
gatcagcagc tctactacaa gtcctacttc ctcttcatct ttggtgtcga gttcgtgggc 1440
cctgtggtca ccatgaccct gtgctatgcc aggatctccc gggagctctg gttcaaggca 1500
gtccctgggt tccagacgga gcagattcgc aagcggctgc gctgccgcag gaagacggtc 1560
ctggtgctca tgtgcattct cacggcctat gtgctgtgct gggcaccctt ctacggtttc 1620
accatcgttc gtgacttctt ccccactgtg ttcgtgaagg aaaagcacta cctcactgcc 1680
ttctacgtgg tcgagtgcat cgccatgagc aacagcatga tcaacaccgt gtgcttcgtg 1740
acggtcaaga acaacaccat gaagtacttc aagaagatga tgctgctgca ctggcgtccc 1800
tcccagcggg ggagcaagtc cagtgctgac cttgacctca gaaccaacgg ggtgcccacc 1860
acagaagagg tggactgtat caggctgaag tgacccactg gtgtcacaca attgaaaacc 1920
ccagtccagt actcagagca tcacccacca tcaaccaagt tcataggctg catgggaaat 1980
gacatctgtg ttcatgcctc ccccgtgccc tcaagaagcc gaatgctgca aagtcgtaac 2040
atacaatgag actagacatg aaccaaatca gctgacattt actgatatcc gctcgacacc 2100
tactgtgtcc acaatcccaa caaggagatt agacacaagg agcagcaact gacatggact 2160
gaacatgtac tgtgtgcaag ccaaaccaat gagattaaca gggacagcag gagctgaatt 2220
atcttactat gtatcaaacc tgttgttcac aaattaaact acagtccaac ttgggtcaca 2280
tcgttttatt tcccattcat tttt 2304

<210> 32
<211> 2322
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482007CB1

<400> 32
cccatttcaa aaatggagaa gacagatcac tgccactgac caggaccgtg ggaggtgcca 60
cgtgatggtg aggcatcatg ctagggagct gagctctgac cttcctgctg ggtgattctc 120
cacctctggg ctgctagatc tacttcctgg atgccgtgaa gatcctcatg tatgaaaatg 180
aagtcccagg caaccatgat ttgctgctta gtgttctttc tgtccacaga atgttcccac 240
tatagatcca agattcacct aaaagctgga gataaacttc aaagccctga agggaaaccc 300
aagactggaa ggatccaaga gaaatgcgaa ggaccttgta tttcttcttc caactgcagc 360
cagccctgtg ctaaggactt tcatggagaa ataggattta catgtaatca aaaaaagtgg 420
caaaaatcag ctgaaacatg tacaagcctt tctgtggaaa aactctttaa ggactcaact 480
ggtgcatctc gcctttctgt agcagcacca tctatacctc tgcatattct agactttcga 540
gctccagaga ccattgagag tgtagctcaa ggaatccgta agaactgccc ctttgattat 600
gcctgcatca ctgacatggt gaaatcatca gaaacaacat ctggaaatat tgcatttata 660
gtggagttat taaaaaatat ttctacagac ttgtctgata atgttactcg agagaaaatg 720
aagagctata gtgaagtggc caaccacatc ctcgacacag cagccatttc aaactgggct 780
ttcattccca acaaaaatgc cagctcggat ttgttgcagt cagtgaattt gtttgccaga 840
caactccaca tccacaataa ttctgagaac attgtgaatg aactcttcat tcagacaaaa 900
gggtttcaca tcaaccataa tacctcagag aaaagcctca atttctccat gagcatgaac 960
aataccacag aagatatctt aggaatggta cagattccca ggcaagagct aaggaagctg 1020
tggccaaatg catcccaagc cattagcata gctttcccaa ccttgggggc tatcctgaga 1080
gaagcccact tgcaaaatgt gagtcttccc agacaggtaa atggtctggt gctatcagtg 1140
gttttaccag aaaggttgca agaaatcata ctcaccttcg aaaagatcaa taaaacccgc 1200
aatgccagag cccagtgtgt tggctggcac tccaagaaaa ggagatggga tgagaaagcg 1260
tgccaaatga tgttggatat caggaacgaa gtgaaatgcc gctgtaacta caccagtgtg 1320
gtgatgtctt tttccattct catgtcctcc aaatcgatga ccgacaaagt tctggactac 1380
atcacctgca ttgggctcag cgtctcaatc ctaagcttgg ttctttgcct gatcattgaa 1440
gccacagtgt ggtcccgggt ggttgtgacg gagatatcat acatgcgtca cgtgtgcatc 1500
gtgaatatag cagtgtccct tctgactgcc aatgtgtggt ttatcatagg ctctcacttt 1560
aacattaagg cccaggacta caacatgtgt gttgcagtga cattttcag ccacttttc 1620
tacctctctc tgtttttctg gattctcttc aaagcattgc tcatcattta tggaatattg 1680
gtcattttcc gtaggatgat gaagtcccga atgatggtca ttggctttgc cattggctat 1740
gggtgcccat tgatcattgc tgtcactaca gttgctatca cagggccagt gaaaggctac 1800
atgagacctg aggcctgttg gcttaactgg gacaatacca aagccctttt agcatttgcc 1860
atcccggcgt tcgtcattgt ggctgtaaat ctgattgtgg ttttggttgt tgctgtcaac 1920
actcagaggc cctctattgg cagttccaag tctcaggatg tggtcataat tatgaggatc 1980
agcaaaaatg ttgccatcct cactccactg ctgggactga cctggggttt tggaatagcc 2040

actctcatag aaggcacttc cttgacgttc catataattt ttgccttgct caatgctttc 2100
cagggttttt tcatcctgct gtttggaacc attatggatc acaagataag agatgctttg 2160
aggatgagga tgtcttcact gaaggggaaa tcgagggcag ctgagaatgc atcactaggc 2220
ccaaccaatg gatctaaatt aatgaatcgt caaggatgaa atgctgcccc atttctcatg 2280
gatgtcctga gaccaagagg ggagatccag gagaaagagg cc 2322

<210> 33
<211> 2366
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6769042CB1

<400> 33
atttaggtga cactatagaa gagcccagtg tgctggaaag gagatcgcca tgtacttcac 60
tgctgccatt ggaaagcatg ctttattgtc ttcaacgctg ccaagcctct tcatgacatc 120
cacagcaagc cccgtgatgc ccacagatgc ctaccatccc atcataacca acctgacaga 180
agagagaaaa accttccaaa gtcccggagt gatactgagt tacctccaaa atgtatccct 240
cagcttaccc agtaagtccc tctcggagca gacagccttg aatctcacca agaccttctt 300
aaaagccgtg ggagagatcc ttctactgcc tggttggatt gctctgtcag aggacagcgc 360
cgtggtactg agtctcatcg acactattga caccgtcatg ggccatgtat cctccaacct 420
gcacggcagc acgccccagg tcaccgtgga gggctcctct gccatggcag agttttccgt 480
ggccaaaatc ctgcccaaga ccgtgaattc ctcccattac cgcttcccgg cccacgggca 540
gagcttcatc cagatccccc acgaggcctt ccacaggcac gcctggagca ccgtcgtggg 600
tctgctgtac cacagcatgc actactacct gaacaacatc tggcccgccc acaccaagat 660
cgcggaggcc atgcatcacc aggactgcct gctgttcgcc accagccacc tgatttccct 720
ggaggtgtcc ccaccaccca ccctgtctca gaacctgtcg ggctctccac tcattacggt 780
ccacctcaag cacagattga cacgtaagca gcacagtgag gccaccaaca gcagcaaccg 840
agtcttcgtg tactgcgcct tcctggactt cagctccgga gaaggggtct ggtcgaacca 900
cggctgtgcg ctcacgagag gaaacctcac ctactccgtc tgccgctgca ctcacctcac 960
caactttgcc atcctcatgc aggtggtccc gctggagctt gcacgcggac accaggtggc 1020
gctgtcgtct atcagctatg tgggctgctc cctctccgtg ctctgcctgg tggccacgct 1080
ggtcaccttc gccgtgctgt cctccgtgag caccatccgg aaccagcgct accacatcca 1140
cgccaacctg tccttcgccg tgctggtggc ccaggtcctg ctgctcatta gtttccgcct 1200
cgagccaggc acgacccccc gccaagtgat ggccgtgctc ctacactact tcttcctgag 1260
tgccttcgca tggatgctgg tggaggggct gcacctctac agcatggtga tcaaggtctt 1320
tgggtcggag gacagcaagc accgttacta ctatgggatg ggatggggtt ttcctcttct 1380
gatctgcatc atttcactgt catttgccat ggacagttac ggaacaagca acaattgctg 1440
gctgtcgttg gcgagtggcg ccatctgggc ctttgtagcc cctgccctgt ttgtcatcgt 1500
ggtcaacatt ggcatcctca tcgctgtgac cagagtcatc tcacagatca gcgccgacaa 1560
ctacaagatc catggagacc ccagtgcctt caagttgacg gccaaggcag tggccgtgct 1620
gctgcccatc ctgggtacct cgtgggtctt tggcgtgctt gctgtcaacg gttgtgctgt 1680
ggttttccag tacatgtttg ccacgctcaa ctccctgcag ggactgttca tattcctctt 1740
tcattgtctc ctgaattcag aggtgagagc cgccttcaag cacaaaatca aggtctggtc 1800
gctcacgagc agctccgccc gcacctccaa cgcgaagccc ttccactcgg acctcatgaa 1860
tgggacccgg ccaggcatgg cctccaccaa gctcagccct tgggacaaga gcagccactc 1920
tgcccaccgc gtcgacctgt cagccgtgtg agccgggagg ctgccaacca ggccaggctg 1980
cgctcagaac acacccccc aaacagaatg aaatgcccca cctttgccca tggaccctct 2040
ccttgctgct gtctggacat gggtgttgtg gccccgagac agctgtcctc ccctgtgact 2100
ctggctgtcg gagcacactg ctcagcccag cagcctgatg cccaggccag cgtgggccct 2160
cctgccttgc atccacccgt gggctgagtg acttcctcgg gggattccca ggacacagtg 2220
gcctgacttg tgatggtgcc cttgagcctc ccttcatcac tcagcatcag accagcgagg 2280
cagggcatcg gggccggtcc cgcagcccgg agggatgtca gctctgtgct ggggggttgg 2340
ggcccgcccc aagtgtcagg ccccgc 2366

<210> 34
<211> 1458
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 7476053CB1

<400> 34
atggaggccg ctagcctttc agtggccacc gccggcgttg cccttgccct gggacccgag 60
accagcagcg ggaccccaag cccgagaggg atactcggtt cgaccccgag cggcgccgtc 120
ctgccgggcc gagggccgcc cttctctgtc ttcacggtcc tggtggtgac gctgctagtg 180
ctgctgatcg ctgccacttt cctgtggaac ctgctggttc cggtcaccat cccgcgggtc 240
cgtgccttcc accgcgtgcc gcataacttg gtggcctcga cggccgtctc ggacgaacta 300
gtggcagcgc tggcgatgcc accgagcctg gcgagtgagc tgtcgaccgg gcgacgtcgg 360
ctgctgggcc ggagcctgtg ccacgtgtgg atctccttcg acgccctgtg ctgccccgcc 420
ggcctcggga acgtggcggc catcgccctg ggccgcgacg gggccatcac acggcacctg 480
cagcacacgc tgcgcacccg cagccgcgcc tcgttgctca tgatcgcgct cgcccgggtg 540
ccgtcggcgc tcatcgccct cgcgccgctg ctctttggcc ggggcgaggt gtgcgacgct 600
cggctccagc gctgccaggt gagccgggaa ccctcctatg ccgccttctc cacccgcggc 660
gccttccacc tgccgcttgg cgtggtgccg tttgtctacc ggaagatcta cgaggcggcc 720
aagtttcgtt tcggccgccg ccggagagct gtgctgccgt tgccggccac catgcaggtg 780
aaggaagcac ctgatgaggc tgaagtggtg ttcacggcac attgcaaagc aacggtgtcc 840
ttccaggtga gcggggactc ctggcgggag cagaaggaga ggcgagcagc catgatggtg 900
ggaattctga ttggcgtgtt tgtgctgtgc tggatcccct tcttcctgac ggaactcatc 960
agcccactct gtgcctgcag cctgcccccc atctggaaaa gcatatttct gtggcttggc 1020
tactccaatt ctttcttcaa ccccctgatt tacacagctt ttaacaagaa ctacaacaat 1080
gccttcaaga gcctctttac taagcagaga tgaacacagg ggttagagag acatgggtag 1140
attttaagga ggaaggaact tggacttttt cgtcagtgat ctgagattct tccctccaca 1200
gctgagtgct aatgctgtat tgagagttat accattgggc ctggactgta gaagcagcag 1260
agccaaggtt ctcaagaaag acagcaaagg tctggcagat gttgtaacta tgccttcttc 1320
ccatgtgcat ggcagacatt gccaattggt catggcttgg ctccccactg agcaggaact 1380
tggtctcaga atcctttcca ggacagcacc ctaggcagct actgttgatt atttaaaatt 1440
gatgcaagac ttgaaaaa 1458

<210> 35
<211> 975
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480410CB1

<400> 35
atgggcatgg agggtcttct ccagaactcc actaacttcg tcctcacagg cctcatcacc 60
catcctgcct tccccgggct tctctttgca atagtcttct ccatctttgt ggtggctata 120
acagccaact tggtcatgat tctgctcatc cacatggact cccgcctcca cactcccatg 180
tacttcttgc tcagccagct ctccatcatg gataccatct acatctgtat cactgtcccc 240
aagatgctcc aggacctcct gtccaaggac aagaccattt ccttcctggg ctgtgcagtt 300
cagatcttcc tctacctgac cctgattgga ggggaattct tcctgctggg tctcatggcc 360
tatgaccgct atgtggctgt gtgcaaccct ctacggtacc ctctcctcat gaaccgcagg 420
gtttgcttat tcatggtggt cggctcctgg gttggtggtt ccttggatgg gttcatgctg 480
actcctgtca ctatgagttt ccccttctgt agatcccgag agatcaatca ctttttctgt 540
gagatcccag ccgtgctgaa gttgtcttgc acagacacgt cactctatga gaccctgatg 600
tatgcctgct gcgtgctgat gctgcttatc cctctatctg tcatctctgt ctcctacacg 660
cacatcctcc tgactgtcca caggatgaac tctgctgagg gccggcgcaa agcctttgct 720
acgtgttcct cccacattat ggtggtgagc gttttctacg gggcagcctt ctacaccaac 780
gtgctgcccc actcctacca cactccagag aaagataaag tggtgtctgc cttctacacc 840
atcctcaccc ccatgctcaa cccactcatc tacagcttga ggaataaaga tgtggctgca 900
gctctgagga aagtactagg gagatgtggt tcctcccaga gcatcagggt ggcgactgtg 960
atcaggaagg gctag 975

<210> 36
<211> 948
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55036418CB1

<400> 36
atggagacgt gggtgaacca gtcctacaca gatggcttct tcctcttagg catcttctcc 60
cacagtactg ctgaccttgt cctcttctcc gtggttatgg cggtcttcac agtggccctc 120
tgtgggaatg tcctcctcat cttcctcatc tacatggacc ctcaccttca cacccccatg 180
tacttcttcc tcagccagct ctccctcatg gacctcatgt tggtctgtac caatgtgcca 240
aagatggcag ccaacttcct gtctggcagg aagtccatct cctttgtggg ctgtggcata 300
caaattggcc tctttgtctg tcttgtggga tctgaggggc tcttgctggg actcatggct 360
tatgaccgct atgtggccat tagccaccca cttcactatc ccatcctcat gaatcagagg 420
gtctgtctcc agattactgg gagctcctgg gcctttggga taatcgatgg cttgatccag 480
atggtggtag taatgaattt cccctactgt ggcttgagga aggtgaacca tttcttctgt 540
gagatgctat ccttgttgaa gctggcctgt gtagacacat ccctgtttga gaaggtgata 600
tttgcttgct gtgtcttcat gcttctcttc ccattctcca tcatcgtggc ctcctatgct 660
cacattctag ggactgtgct gcaaatgcac tctgctcagg cctggaaaaa ggccctggcc 720
acctgctcct cccacctgac agctgtcacc ctcttctatg gggcagccat gttcatctac 780
ctgaggccta ggcactaccg ggcccccagc catgacaagg tggcctctat cttctacacg 840
gtccttactc ccatgctcaa ccccctcatt tacagcttga ggaacaggga ggtgatgggg 900
gcactgagga aggggctgga ccgctgcagg atcggcagcc agcactga 948

<210> 37
<211> 1086
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481701CB1

<400> 37
ggctctattc agacgctggc ttcttgtaag tgtattcctt tatccaatag tagatgcctc 60
ctagaaggct tgagtgcact ggaaattgaa actctcactt tcaacttgga gatggagagc 120
cccaatcaaa ccaccattca ggagtttatc ttctccgctt tcccttattc ctgggttaag 180
tctgttgtct gctttgttcc actgctcttc atctatgctt tcattgttgt tggaaacctg 240
gtcatcatca cagtggtcca gttgaatact cacctccaca ctcccatgta tacttttatc 300
agtgctcttt cttttctgga gatttggtat accacagcca caatcccaaa gatgctgtct 360
agcctgctta gtgagaggag catttccttc aatggttgtc tcctgcagat gtatttcttc 420
cattccaccg gcatctgtga ggtgtgtctc ttgacagtta tggcctttga ccactacctg 480
gccatatgca gccctcttca ttatccctct atcatgaccc ccaagctatg tacccaactg 540
actttaagtt gctgtgtttg tggctttatc acacccgttc ctgagattgc ctggatctct 600
acactgccat tttgtggttc gaatcacctt gaacatatct tctgtgactt cctcccagtg 660
ctgcgtctgg cctgcacaga cacacgagcc atcgtcatga ttcaggtagt ggatgtcatt 720
catgcagtgg agattattac agctgtgatg ctcatcttca tgtcctacga tggtattgtg 780
gctgtaattc tacgtattca ttcagctgga ggccgccgca cagcattttc cacgtgtgtc 840
tctcacttca ttgtcttttc gctcttcttt ggcagtgtga ctctcatgta cctacgcttc 900
tctgccacct actctttgtt ctgggatata gccattgctc tggcctttgc agttttgtct 960
cccttcttca accccattat ctatagcctg aggaataaag aaataaaaga agctataaaa 1020
aagcacatag gtcaagctaa gatatttttt tccgtaagac cagggacctc aagtaagata 1080
ttttag 1086

<210> 38
<211> 1529
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481774CB1

<400> 38
aagggagacc acagtgagag ggagccctga gcagaagtaa ggctgtcaca aggctggaag 60

cagagaacat ccccatggaa ctgaagacag catgctgcat ccctgggagg agggagctct 120
taaggaagtt ccaaggattg atatttctgt tcagctgcag tagagatgga tggaaccatg 180
gcagcaccca acccatttca tcctactggg attctctgac cgaccccatc tggagaggat 240
cctctttgtg gtcatcctga tcgcgtacct cctgaccctc gtaggcaaca ccaccatcat 300
cctggtgtcc cggctggacc cccacctcca cacccccatg tacttcttcc tcgcccacct 360
ttccttcctg gacctcagtt tcaccaccag ctccatcccc cagctgctct acaaccttaa 420
tggatgtgac aagaccatca gctacatggg ctgtgccatc cagctcttcc tgttcctggg 480
tctgggtggt gtggagtgcc tgcttctggc tgtcatggcc tatgaccggt gtgtggctat 540
ctgcaagccc ctgcactaca tggtgatcat gaaccccagg ctctgccggg gcttggtgtc 600
agtgacctgg ggctgtgggg tggccaactc cttggccatg tctcctgtga ccctgcgctt 660
accccgctgt gggcaccacg aggtggacca cttcctgtgt gagatgcccg ccctgatccg 720
gatggcctgc atcagcactg tggccatcga cggcaccgtc tttgtcctgg cggtgggtgt 780
tgtgctgtcc cccttggtgt ttatcctgct ctcttacagc tacattgtga gggctgtgtt 840
acaaattcgg tcagcatcag gaaggcagaa ggccttcggc acctgcggct cccatctcac 900
tgtggtctcc cttttctatg gaaacatcat ctacatgtac atgcagccag gagccagttc 960
ttcccaggac cagggcaagt tcctcacgct cttctacaac attgtcaccc ccctcctcaa 1020
tcctctcatc tacaccctca gaaacagaga ggtgaagggg gcactgggaa ggttgcttct 1080
ggggaagaga gagctaggaa aggagtaaag gcatctccac ctgacttcac ctccatccag 1140
ggccactggc agcatctgga acggctgaat tccagctgat attagcccac gactcccaac 1200
ttgccttttt ctggactttt gtgaggctgt ttcagttctg acattatgtg tttttgttgt 1260
tgctcttaaa attgagacgg ggtctcactc tgtcacctag ggtggagtgc agtggtgcca 1320
ccatagctcc ttcgactatt gggcttaagc gatcctcccc cacctcagcc ttccaagtaa 1380
ctgggactac aggtgtgcat cactggcagt gggaattgtg gcttttctgt cttctatgga 1440
gacggggtct tgctgtgttg accaggctgg tcccaaactc ctggcctcat gtgatcctcc 1500
tgccatggcc tcctaaagtt ctgggatta 1529